

# STEIN'S METHOD OF EXCHANGEABLE PAIRS FOR THE BETA DISTRIBUTION AND GENERALIZATIONS

CHRISTIAN DÖBLER

**ABSTRACT.** We propose a new version of Stein's method of exchangeable pairs, which, given a suitable exchangeable pair  $(W, W')$  of real-valued random variables, suggests the approximation of the law of  $W$  by a suitable absolutely continuous distribution. This distribution is characterized by a first order linear differential Stein operator, whose coefficients  $\gamma$  and  $\eta$  are motivated by two regression properties satisfied by the pair  $(W, W')$ . Furthermore, the general theory of Stein's method for such an absolutely continuous distribution is developed and a general characterization result as well as general bounds on the solution to the Stein equation are given. This abstract approach is a certain extension of the theory developed in the papers [5] and [13], which only consider the framework of the density approach, i.e.  $\eta \equiv 1$ . As an illustration of our technique we prove a general plug-in result, which bounds a certain distance of the distribution of a given random variable  $W$  to a Beta distribution in terms of a given exchangeable pair  $(W, W')$  and provide new bounds on the solution to the Stein equation for the Beta distribution, which complement the existing bounds from [18]. The abstract plug-in result is then applied to derive bounds of order  $n^{-1}$  for the distance between the distribution of the relative number of drawn red balls after  $n$  drawings in a Pólya urn model and the limiting Beta distribution measured by a certain class of smooth test functions.

## 1. INTRODUCTION

Since its introduction in [31] in 1972 Stein's method has become a famous and useful tool for proving distributional convergence. One of its main advantages over other techniques is that it automatically yields concrete error bounds on various distributional distances. Being first only developed for normal approximation it was observed by several authors that Stein's idea of linking a characterizing operator for the target distribution to a differential equation, the *Stein equation*, carries over to many other absolutely continuous and discrete distributions, where, in the discrete case, the differential equation has to be replaced by a suitable difference equation. Among those other distributions, to which Stein's method has been successfully extended, are the Poisson distribution (see e.g. [6], [1] or [3]), the Gamma distribution (see [22] or [28]), the exponential distribution (see e.g. [4], [25] and [14]), the Laplace distribution [27] and, more generally, the class of Variance-Gamma distributions [16]. Stein's method for the Beta distribution has been developed independently in the paper [18] as well as in the preprint [11].

Although in both works [18] and [11] a rate of convergence for the relative number of drawn red balls in a Pólya urn model was derived using Stein's method for the Beta distribution, the actual approaches were quite different. In [18] the authors

---

Université du Luxembourg, Unité de Recherche en Mathématiques  
christian.doebler@uni.lu

*Keywords:* Stein's method, exchangeable pairs, Beta distribution, Pólya urn model.

developed a useful and widely applicable technique to find a whole class of characterizing operators for a discrete distribution, whose probability mass function is known explicitly, and compared one of these operators to the Stein operator of the limiting Beta distribution. In contrast, the preprint [11] built on a coupling approach by developing a new version of the *exchangeable pairs approach* of Stein's method for a rather large class of absolutely continuous distributions on the real line. This new version of the exchangeable pairs approach differs from that in the framework of the *density method* as developed in [13] and [5], since it allows for a modification of the Stein equation, which is adapted to a given exchangeable pair and does not necessarily rely on the characterization by the density method. Recently, in [21], a nice generalization of the density method, which does not necessarily assume absolute continuity of the given distribution, was given and, as an application, it was shown, how the situation of the Pólya urn example from [18] may be fitted into this framework.

The main purpose of the present paper is to give a more easily readable account of the method and ideas from [11] by keeping the class of Beta distributions on  $[0, 1]$  and the Pólya urn model as a running example. In addition, we derive new numerical bounds on the solution to the Stein equation for the Beta distribution and, for smooth test functions, also on its first order derivative. For Lipschitz-continuous test functions, these bounds complement those given in [18] in the sense that they are neither uniformly worse nor uniformly better in the parameters of the Beta distribution. Furthermore, we use a new iterative procedure to obtain uniform bounds for derivatives of any order of the solution to the Beta Stein equation with sufficiently smooth right hand side. Incidentally, this is the first paper to give bounds on higher order derivatives of the solution to the Beta Stein equation. It should be mentioned that, generally, obtaining bounds on higher order derivatives of the solution to the Stein equation is quite a difficult problem, because the explicit representations of those derivatives become more and more complicated. Hence, to date bounds on higher order derivatives of the solution are still quite rare in Stein's method. For instance, the paper [9] obtains sharp bounds on higher order derivatives in the context of the normal and exponential distributions by exploiting very peculiar identities and facts about these distributions, which are not available for more general absolutely continuous distributions. Also, if one succeeds in deriving a tractable generator representation of the solution to the Stein equation as suggested in [2], one can usually use this form of the solution to obtain bounds on higher order derivatives. This has been used for the multivariate normal [19] and for the Gamma distribution [22]. However, in contrast to the bounds from [9], these bounds usually do not exhibit the smoothness property of the inverse of the corresponding Stein operator. In the case of the multivariate normal distribution with non-singular covariance matrix, one can combine the generator representation with a partial integration to obtain bounds on higher order derivatives, which demand one fewer order of smoothness from the test function than the bounds from [19]. This has been accomplished independently in [15] and [10]. The recent paper [17] combines bounds obtained from the generator representation with the iterative method from the present article in order to obtain new bounds on derivatives of arbitrary order of the solution to the Gamma Stein equation, whose dependence on the shape parameter of the Gamma distribution is superior to previous bounds.

We also indicate, how our iterative method can be applied to obtain bounds for the solution to a Stein equation for the exponential distribution, which are better than those previously obtained. We thus suggest that exploiting this iterative procedure can become a fruitful technique for a larger class of distributions.

The remainder of this paper is structured as follows: In Section 2 the general approach is motivated by means of a natural exchangeable pair in the context of the Pólya urn model and it is stressed by means of this example that the framework of exchangeable pairs within the density approach as developed in [13] and [5] is not always suitable and why one might want to use a different Stein characterization. Furthermore, our main application, Theorem 2.1, a quantitative distributional limit theorem for the relative number of drawn red balls is stated. Then, motivated by this example, in Section 3 a general version of Stein's method for a large class of absolutely continuous distributions adapted to a given exchangeable pair is developed. In Section 4 the theory from Section 3 is specialized to the class of Beta distributions and Theorem 2.1 is proved. Finally, in Section 5 several proofs for statements from Sections 3 and 4 are given.

#### ACKNOWLEDGEMENTS

Most parts of the research which led to this article have been accomplished during the authors PhD studies and, hence, there is a certain overlap with the author's PhD thesis [10], see also the unpublished paper [11]. The author was supported by the DFG via SFB/TR 12 during this time. We also refer to Appendix A of [10] for a version of de l'Hôpital's rule which covers (locally) absolutely continuous functions and which is general enough to justify all invocations of this famous tool within this article. Finally, Appendix B of [10] contains some identities about the Gibbs sampling procedure, which are generally useful in the exchangeable pairs version of Stein's method and which will be used in the present paper. I am grateful to an anonymous referee whose detailed and valuable comments and suggestions helped me improve the presentation of my results.

## 2. THE PÓLYA URN MODEL AND MOTIVATION OF OUR GENERAL APPROACH

The classical Pólya urn model can be described as follows. Fix positive integers  $r, w$  and  $c$ . At the beginning an urn contains  $r$  red balls and  $w$  white balls. At each discrete time point  $n \in \mathbb{N}$  a ball is drawn from the urn uniformly at random and this ball together with  $c$  other balls of the same colour is returned to the urn. If we denote by  $S_n$  the number of drawn red balls after the first  $n$  drawings,  $n \in \mathbb{N}$ , then we can write

$$(1) \quad S_n = \sum_{j=1}^n X_j,$$

where  $X_j$  denotes the indicator of the event that the  $j$ -th drawn ball is red,  $j \in \mathbb{N}$ . It is known from elementary probability theory that for each  $n \in \mathbb{N}$  and all  $x_1, \dots, x_n \in \{0, 1\}$  we have

$$(2) \quad P(X_1 = x_1, \dots, X_n = x_n) = \frac{\prod_{i=0}^{k-1} (r + ci) \prod_{j=0}^{n-k-1} (w + cj)}{\prod_{l=0}^{n-1} (s + w + cl)},$$

where  $k := \sum_{j=1}^n x_j$ . In particular, this shows that the sequence  $(X_j)_{j \in \mathbb{N}}$  is exchangeable. It now follows from (2) that for each  $k = 0, \dots, n$  we have

$$P(S_n = k) = \binom{n}{k} \frac{\prod_{i=0}^{k-1} (r + ci) \prod_{j=0}^{n-k-1} (w + cj)}{\prod_{l=0}^{n-1} (s + w + cl)},$$

or, with  $a := \frac{r}{c}$  and  $b := \frac{w}{c}$ ,

$$(3) \quad P(S_n = k) = \frac{\binom{-a}{k} \binom{-b}{n-k}}{\binom{-a-b}{n}},$$

where, for a real number  $x$  and a nonnegative integer  $m$ , we define the generalized binomial coefficient by

$$\binom{x}{m} := \frac{x(x-1) \cdots (x-m+1)}{m!}.$$

The distribution of  $S_n$  given by (3) is usually referred to as the *Pólya distribution* with parameters  $n \in \mathbb{N}$  and  $a, b > 0$ . It is a well-known fact that the distribution of  $\frac{1}{n}S_n$  converges weakly as  $n \rightarrow \infty$  to the distribution  $Beta(a, b)$  with parameters  $a$  and  $b$ , where, for general  $a, b > 0$ , the Beta distribution  $Beta(a, b)$  with parameters  $a$  and  $b$  is defined by the density function  $p := p_{a,b}$  with

$$(4) \quad p_{a,b}(x) := \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{else.} \end{cases}$$

Here,  $B(a, b)$  denotes the Euler Beta function  $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$  which is related to the Gamma function  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$  via

$$(5) \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

From now on denote by

$$(6) \quad W := W_n := \frac{1}{n}S_n = \frac{1}{n} \sum_{j=1}^n X_j$$

the relative number of drawn red balls after the first  $n$  drawings from the urn. Denote by  $C^{1,1}([0, 1]; \mathbb{R})$  the space of all continuously differentiable real-valued functions on  $[0, 1]$  which have a Lipschitz-continuous derivative.

**THEOREM 2.1.** *Let  $Z \sim Beta(a, b)$ . For each  $h \in C^{1,1}([0, 1]; \mathbb{R})$  we have that*

$$\begin{aligned} & \left| E[h(W)] - E[h(Z)] \right| \\ & \leq \frac{C(a, b)}{n} \|h'\|_\infty \left( \frac{ab}{a+b} + \frac{(a+b)C(a+1, b+1)}{6} \left( 1 + \frac{a+b-1}{n} \right) \right) \\ & \quad + \frac{C(a+1, b+1)}{6n} \|h''\|_\infty \left( 1 + \frac{a+b-1}{n} \right), \end{aligned}$$

where the constants  $C(\cdot, \cdot)$  are defined in (47) and (48) below and  $\|h''\|_\infty$  denotes the minimum Lipschitz constant of  $h'$ .

The proof will be given in Section 4. In the paper [18] the authors even proved a concrete upper bound of order  $n^{-1}$  for the Wasserstein distance between the distributions of  $Z$  and  $W$  from Theorem 2.1 and also showed that the rate  $n^{-1}$  is optimal. Since the Wasserstein distance is induced by 1-Lipschitz test functions, this implies that their result is stronger than Theorem 2.1 as far as the class of test functions is concerned. However, it should be mentioned that their method of comparing Stein operators can only be applied in situations, where the distribution of  $W$  is explicitly known. Contrarily, the exchangeable pairs technique which is used here, in general, seems to be more flexible in this respect. For instance, our plug-in result, Theorem 4.4 below, might be beneficial for other applications, where the exact distribution of  $W$  is not at hand. Moreover, even in the situation of Theorem 2.1 there exist parameters  $a, b > 0$  and test functions  $h$  such that our bound is smaller than the one obtained in [18]. To see this, fix  $n$  and let  $a = b$  tend to zero. Also, let  $h \in C^{1,1}([0, 1]; \mathbb{R})$  be such that  $\|h'\|_\infty = \|h''\|_\infty = 1$ . Then, as  $C(a, a) = 4$  if  $a \leq 1$  and by continuity of  $C(a, a)$  in  $a$ , we see that the bound given in Theorem 2.1 converges to  $\frac{2}{3n}(1 - 1/n) \leq \frac{2}{3n}$ , whereas the bound from [18] converges to the bigger value  $\frac{9}{2n}$ .

Recall that a pair  $(X, X')$  of random elements on a common probability space is called exchangeable, if

$$(X, X') \stackrel{\mathcal{D}}{=} (X', X).$$

Representation (6) for  $W$  suggests constructing another random variable  $W'$  such that  $W$  and  $W'$  make up an exchangeable pair using a Gibbs sampling procedure. Noticing that also the random variables  $X_1, \dots, X_n$  are exchangeable, the construction of  $W'$  can be simplified to the following:

Observe  $X_1 = x_1, \dots, X_n = x_n$  and construct  $X'_n$  according to the distribution  $\mathcal{L}(X_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1})$ . Then, letting

$$(7) \quad W' := W - \frac{1}{n}X_n + \frac{1}{n}X'_n$$

the pair  $(W, W')$  is exchangeable. Note that  $|W - W'| \leq \frac{1}{n}$  is small which suggests that the exchangeable pair  $(W, W')$  be beneficial for a Stein's method approach to the proof of weak convergence of  $\mathcal{L}(W_n)$  to  $Beta(a, b)$ . From the exchangeable pairs approach within normal approximation (see e.g. [32], [8] or [7]) and for non-normal approximation (see [13] and [5]) we know that exchangeability of  $(W, W')$  is not enough to guarantee distributional closeness of  $W$  and of  $Z \sim Beta(a, b)$  but that a further *regression property* has to be satisfied.

**PROPOSITION 2.2.** *The exchangeable pair  $(W, W')$  satisfies the regression property*

$$E[W' - W | W] = \frac{a + b}{n(a + b + n - 1)} \left( \frac{a}{a + b} - W \right) = \lambda \gamma_{a,b}(W),$$

where  $\gamma_{a,b}(x) = (a + b) \left( \frac{a}{a + b} - x \right)$  and  $\lambda = \lambda_n = \frac{1}{n(a + b + n - 1)}$ .

*Proof.* We have  $W' - W = \frac{X'_n}{n} - \frac{X_n}{n}$  and by exchangeability of  $X_1, \dots, X_n$  it clearly holds that  $E[X_n | W] = E[X_n | S_n] = \frac{1}{n}S_n = W$ . Also, by the definition of  $X'_n$  and

since  $X'_n$  only assumes the values 0 and 1 we have for any  $x_1, \dots, x_{n-1} \in \{0, 1\}$

$$\begin{aligned} E[X'_n | X_1 = x_1, \dots, X_n = x_n] &= E[X_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}] \\ &= P(X_n = 1 | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = \frac{r + c \sum_{j=1}^{n-1} x_j}{r + w + c(n-1)}, \end{aligned}$$

and hence,

$$E[X'_n | X_1, \dots, X_n] = \frac{r + c \sum_{j=1}^{n-1} X_j}{r + w + c(n-1)} = \frac{r + cnW - cX_n}{r + w + c(n-1)}.$$

Thus, since  $\sigma(W) \subseteq \sigma(X_1, \dots, X_n)$ , we obtain

$$\begin{aligned} E[X'_n | W] &= E[E[X'_n | X_1, \dots, X_n] | W] = \frac{r + cnW - cW}{r + w + c(n-1)} \\ &= \frac{r + c(n-1)W}{r + w + c(n-1)} = \frac{a + (n-1)W}{a + b + n - 1}. \end{aligned}$$

Finally, we have

$$\begin{aligned} E[W' - W | W] &= \frac{1}{n} E[X'_n - X_n | W] = \frac{1}{n} \frac{a + (n-1)W}{a + b + n - 1} - \frac{1}{n} W \\ &= \frac{a - (a+b)W}{n(a+b+n-1)} = \frac{a+b}{n(a+b+n-1)} \left( \frac{a}{a+b} - W \right), \end{aligned}$$

as was to be shown. □

From the theory developed in [13] and in [5] we know that if a given exchangeable pair  $(W, W')$  satisfies a regression property of the form

$$(8) \quad \frac{1}{\lambda} E[W' - W | W] = \psi(W) + R,$$

where  $\lambda > 0$  is a typically small constant and  $R$  is negligible in size, then  $\mathcal{L}(W)$  can be approximated by the absolutely continuous distribution whose density has logarithmic derivative  $\psi$ , if and only if the following additional condition is satisfied: It must be the case that

$$(9) \quad \frac{1}{2\lambda} E[(W' - W)^2 | W] \approx 1,$$

which is often paraphrased as that the term on the left hand side in (9) must satisfy a law of large numbers in order for the approximation to be accurate. Comparing (8) to the statement of Proposition 2.2 we see that according to the theory from [13] or [5] the only possibility would be to approximate the distribution of  $W$  by a distribution whose density has logarithmic derivative equal to (a constant multiple) of

$$\frac{a}{a+b} - x,$$

for  $x$  in the support of this density, which should be equal to  $[0, 1]$  in this case. Since the logarithmic derivative  $\psi_{a,b}$  of the density  $p_{a,b}$  of  $Beta(a, b)$  is given by

$$(10) \quad \psi_{a,b}(x) = \frac{d}{dx} \log p_{a,b}(x) = \frac{p'_{a,b}(x)}{p_{a,b}(x)} = \frac{a-1-(a+b-2)x}{x(1-x)}, \quad 0 < x < 1,$$

and we already know that

$$W_n \xrightarrow{\mathcal{D}} \text{Beta}(a, b) \quad \text{as } n \rightarrow \infty,$$

we conclude by way of contradiction that the law of large numbers (9) cannot hold. Indeed, we will see in Proposition 2.3 below that the term on the left hand side of (9) is close to the non-constant random quantity  $W(1 - W)$  rather than to the constant 1. From Proposition 2.2 and some experience with the exchangeable pairs approach within Stein's method we conclude that it would be desirable to have a Stein operator  $L$  of the form

$$(11) \quad Lg(x) = \eta_{a,b}(x)g'(x) + \gamma_{a,b}(x)g(x)$$

for the Beta distribution  $\text{Beta}(a, b)$ . Indeed, in Section 4 we will see that a random variable  $Z \sim \text{Beta}(a, b)$  satisfies the Stein identity

$$(12) \quad E\left[Z(1 - Z)g'(Z) + (a + b)\left(\frac{a}{a + b} - Z\right)g(Z)\right] = 0$$

for all  $g$  in a suitable class of functions, i.e. we can let  $\eta_{a,b}(x) = \eta(x) = x(1 - x)$ . Evidently, the Stein identity (12) was first found in [30] and it was also used in [18]. The statement of the following Proposition will make it possible to exploit the above constructed exchangeable pair  $(W, W')$  in connection with the Stein identity (12) in Section 4.

**PROPOSITION 2.3.** *For the above constructed exchangeable pair  $(W, W')$  we have*

$$E[(W' - W)^2 | W] = \frac{(2n + b - a)W - 2nW^2 + a}{n^2(a + b + n - 1)}$$

and hence

$$\frac{1}{2\lambda} E[(W' - W)^2 | W] = W(1 - W) + \frac{b - a}{2n}W + \frac{a}{2n}.$$

*Proof.* From general facts about Gibbs sampling (see e.g. Appendix B in [10]) it is known that

$$\begin{aligned} E[(W' - W)^2 | W] &= \frac{1}{n^2} \left( E[X_n | W] + E[E[X_n^2 | X_1, \dots, X_{n-1}] | W] \right. \\ &\quad \left. - 2E[X_n E[X_n | X_1, \dots, X_{n-1}] | W] \right). \end{aligned}$$

Since  $X_n^2 = X_n$  we have from the proof of Proposition 2.2 that

$$E[X_n^2 | X_1, \dots, X_{n-1}] = E[X_n | X_1, \dots, X_{n-1}] = \frac{a + nW - X_n}{a + b + n - 1},$$

and hence

$$E[E[X_n^2 | X_1, \dots, X_{n-1}] | W] = \frac{a + (n - 1)W}{a + b + n - 1},$$

where we have used  $E[X_n | W] = W$  again. Finally, we compute

$$\begin{aligned} E[X_n E[X_n | X_1, \dots, X_{n-1}] | W] &= \frac{1}{a + b + n - 1} E[aX_n + nWX_n - X_n^2 | W] \\ &= \frac{aW + nW^2 - W}{a + b + n - 1} = \frac{(a - 1)W + nW^2}{a + b + n - 1}. \end{aligned}$$

Putting pieces together, we eventually obtain

$$(13) \quad \begin{aligned} E[(W' - W)^2 | W] &= \frac{1}{n^2} \left( W + \frac{a + (n-1)W}{a + b + n - 1} - 2 \frac{(a-1)W + nW^2}{a + b + n - 1} \right) \\ &= \frac{(2n + b - a)W - 2nW^2 + a}{n^2(a + b + n - 1)}. \end{aligned}$$

The last assertion easily follows from (13) and from  $\lambda = \frac{1}{n(a+b+n-1)}$ . □

One main aspect of the theoretical contribution of this article is to emphasize that it is no coincidence that

$$\frac{1}{2\lambda} E[(W' - W)^2 | W] \approx \eta(W) = W(1 - W),$$

but that this is a natural replacement of condition (9) from the density approach to our class of Stein operators of the form (15) below.

We end this motivational section by an abstraction of the ideas in the context of the Pólya urn model and the limiting Beta distribution above. Suppose we are given a sequence of random variables  $W = W_n$  of which we know that, as  $n \rightarrow \infty$ , it converges in distribution to a random variable  $Z$  with an absolutely continuous distribution and density  $p$  with respect to the Lebesgue measure. We will also assume that  $p$  itself is absolutely continuous (on each compact subinterval of its support  $(a, b)$ , where  $-\infty \leq a < b \leq \infty$  are extended real numbers). Suppose also that we can naturally construct a random variable  $W'$ , a small random perturbation of  $W$ , such that  $(W, W')$  is an exchangeable pair,  $|W - W'|$  is small in a certain sense and that a regression property of the form

$$(14) \quad \frac{1}{\lambda} E[W' - W | W] = \gamma(W) + R$$

holds, where  $\gamma$  is a certain function on the support of  $\mathcal{L}(Z)$ ,  $\lambda > 0$  is constant and  $R$  is a negligible remainder term. The goal is to compute a rate of convergence for the distributional convergence  $W \rightarrow Z$  by Stein's method of exchangeable pairs for  $\mathcal{L}(Z)$ . By the above reasoning it would be beneficial to have a characterizing Stein operator  $L$  for  $Z$  of the form

$$(15) \quad Lg(x) = \eta(x)g'(x) + \gamma(x)g(x),$$

where  $\eta$  is a function that still has to be found. One might suppose that, in order that  $L$  characterizes  $\mathcal{L}(Z)$ , given the density  $p$  of  $Z$  and the function  $\gamma$  the function  $\eta$  is unique but we will see that this is only so up to a constant multiple of  $p^{-1}$ . Note that by exchangeability

$$(16) \quad 0 = \frac{1}{\lambda} E[W' - W] = E[\gamma(W)] + E[R] \approx E[\gamma(W)] \approx E[\gamma(Z)],$$

where the first approximation is by the assumption that  $R$  is of negligible order and the second is by the fact that  $W$  converges to  $Z$  in distribution. Hence, it is natural to assume from the outset that  $E[\gamma(Z)] = 0$ . In particular, we should assume that  $E|\gamma(Z)| < \infty$ . A natural question is, given  $p$  and  $\gamma$ , if there is a general formula for the function  $\eta$ . In the preprint [11] the first order linear differential equation

$$(17) \quad \eta'(x) = \gamma(x) - \psi(x)\eta(x)$$



was found by making, for a given test function  $h$ , the ansatz  $g_h(x) = \alpha(x)f_h(x)$  for the solutions  $g_h$  of the Stein equation

$$(18) \quad \eta(x)g'(x) + \gamma(x)g(x) = h(x) - E[h(Z)]$$

belonging to the operator (15) and  $f_h$  of the Stein equation

$$(19) \quad f'(x) + \psi(x)f(x) = h(x) - E[h(Z)]$$

corresponding to the density approach. Here, again  $\psi$  denotes the logarithmic derivative of  $p$ . In this paper we follow a different, more direct reasoning. If  $\eta$  is such that (15) is characterizing  $\mathcal{L}(Z)$ , then, for suitable functions  $g$  by partial integration:

$$(20) \quad \begin{aligned} E[\eta(Z)g'(Z)] &= \int_a^b \eta(x)p(x)g'(x)dx = g\eta p \Big|_a^b - \int_a^b (\eta'(x)p(x) + p'(x)\eta(x))g(x)dx \\ &= g\eta p \Big|_a^b - E[(\eta'(Z) + \psi(Z)\eta(Z))g(Z)]. \end{aligned}$$

Thus, if we want this expression to equal

$$g\eta p \Big|_a^b - E[\gamma(Z)g(Z)],$$

then from (20) we conclude that  $\eta$  must satisfy (17). Of course, (17) can be solved by the method of variation of the constant and it turns out that

$$(21) \quad \eta(x) := \frac{1}{p(x)} \int_a^x \gamma(t)p(t)dt, \quad a < x < b,$$

is a particular solution which even satisfies  $(\eta p)(a+) = (\eta p)(b-) = 0$  whenever  $E[\gamma(Z)] = 0$  and, hence, the boundary conditions

$$(22) \quad (g\eta p)(a+) = \lim_{x \downarrow a} g(x)\eta(x)p(x) = 0 = \lim_{x \uparrow b} g(x)\eta(x)p(x) = (g\eta p)(b-),$$

hold for each regular enough, say e.g. bounded, function  $g$ . Also note that every other solution to (17) has the form

$$\eta_\kappa(x) = \eta(x) + \frac{\kappa}{p(x)}$$

for some constant  $\kappa$ . In principle, the particular choice of  $\kappa$  is arbitrary and the choice  $\kappa \neq 0$  sometimes even yields better behaved solutions  $g_h$  to the Stein equation (18). In fact, it is easy to see from (29) below that the choice  $\kappa \neq 0$  automatically implies  $g_h(a+) = g_h(b-) = 0$ . Also, sometimes the choice  $\kappa \neq 0$  is implicit in the density approach. For instance, if  $a > -\infty$ ,  $b < \infty$  and the density  $p$  is such that  $0 \neq p(a+) = p(b-) \in \mathbb{R}$ , then one can easily see that  $\gamma(x) := \psi(x)$  satisfies

$$E[\gamma(Z)] = \int_a^b p'(x)dx = p(b-) - p(a+) = 0$$

but

$$\eta(x) = \frac{p(x) - p(a+)}{p(x)} \neq 1, \quad a < x < b.$$

Hence, in all these cases, using the density approach implicitly entails choosing  $\eta_\kappa$  with  $\kappa = p(a+)$ . When developing the general theory in Section 3 we restrict ourselves to the solution  $\eta$  given by (21), i.e to  $\kappa = 0$ . We thus already mention at this

point that the density approach for  $p$  is included in the theory presented in Section 3 if and only if

$$p(a+) = p(b-) = 0.$$

However, at least if  $\gamma(x) = c(E[Z] - x)$ , it turns out that in many cases  $\eta$  given by (21) has a neat analytical representation, e.g. it is given by a polynomial of degree at most 2, whereas the choice  $\kappa \neq 0$  would introduce a complicated coefficient into (18) originating from the term  $p(x)^{-1}$ . For instance, if  $Z \sim N(0, 1)$  is standard normally distributed and  $\gamma(x) = -x$ , then (21) yields  $\eta \equiv 1$ , whereas the general expression is  $\eta_\kappa(x) = 1 + \kappa e^{x^2/2}$ , which is difficult to handle in practice. Furthermore, if  $p$  is not bounded away from zero, then  $\kappa \neq 0$  gives an unbounded function  $\eta_\kappa$ , whereas  $\eta$  given by (21) usually is bounded, at least if  $a > -\infty$  and  $b < \infty$  (see, e.g. Proposition 3.5 below).

In the next section we will see that under certain mild conditions on the density  $p$  of  $Z$  and on the coefficient  $\gamma$  which, of course, needs not originate from an exchangeable pair, the operator  $L$  given by (15) is indeed characterizing  $\mathcal{L}(Z)$  and prove bounds on the corresponding Stein equation (18) for suitable test functions  $h$ . Finally, we want to propose a strategy of how to proceed, if, contrarily to the above reasoning, we do not know the limiting density  $p$  from the outset but are only given an exchangeable pair  $(W, W')$  such that (14) holds and also

$$(23) \quad \frac{1}{2\lambda} E[(W' - W)^2 | W] = \eta(W) + S$$

is satisfied with the same constant  $\lambda > 0$  and a small remainder term  $S$ , where  $\eta$  now is a certain given function, which is positive on a certain open interval  $J = (a, b) \subseteq \mathbb{R}$ , where  $\gamma$  is also defined. Note that from (17) we have for the logarithmic derivative  $\psi$  of the sought density  $p$  that

$$\psi = \frac{\gamma - \eta'}{\eta}$$

and, hence, for  $x \in J$  and  $x_0 \in J$  an arbitrary point, we have

$$(24) \quad \begin{aligned} p(x) &= p(x_0) \exp\left(\int_{x_0}^x \psi(t) dt\right) = \frac{p(x_0)\eta(x_0)}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) \\ &= \frac{K}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right). \end{aligned}$$

Here, of course,  $K = p(x_0)\eta(x_0)$  is the normalization constant. Formula (24) shows that  $p$  is uniquely determined by  $\gamma$  and  $\eta$ . Furthermore, in Theorem 3.22 we will give precise criteria for  $\gamma$  and  $p$  defined by (24) to satisfy

$$\int_a^b \gamma(t)p(t)dt = 0$$

and for  $\eta$  to satisfy (21) so that the results of the theory developed in Section 3 can in fact be applied. This, together with Proposition 3.19 and Remark 3.20 (iii), suggests the approximation of  $\mathcal{L}(W)$  by the distribution with density  $p$ , if the exchangeable pair  $(W, W')$  satisfies (14) and (23). Note that this idea yields a certain extension of the methodology proposed in [5], where only Stein characterizations from the density approach are put to use.

## 3. THE GENERAL APPROACH

Motivated by Section 2 in this section we develop a general version of Stein's method for a random variable  $Z$  with an absolutely continuous distribution with respect to the Lebesgue measure. This version is useful for those distributions, which allow for a tractable first order linear Stein operator. This class covers many of the standard absolutely continuous distributions. However, it should not be left unmentioned that certain distributions, like the Laplace [27], the Variance-Gamma [16] and the PRR distribution [26] fall outside the scope of this approach, as they only possess a second order linear Stein operator with tractable coefficients.

For an interval  $J \subseteq \mathbb{R}$ , we will call a function defined on  $\mathbb{R}$  *locally absolutely continuous on  $J$* , if its restriction to each compact sub-interval of  $J$  is absolutely continuous. Also, we will use the words increasing, decreasing and so on in the weak sense, unless explicitly otherwise stated. Throughout we suppose that  $Z$  has a Lebesgue density  $p$  on  $\mathbb{R}$  satisfying the following condition:

CONDITION 3.1. For some extended real numbers  $-\infty \leq a < b \leq \infty$  the density  $p$  is positive and locally absolutely continuous on the interval  $(a, b)$ .

By  $\overline{(a, b)}$  we will henceforth denote the closure of the real interval  $(a, b)$  with respect to the usual topology on  $\mathbb{R}$ . Furthermore, we assume that we are given a function  $\gamma$  on  $\overline{(a, b)}$  which might be motivated by a given exchangeable pair and which has the following properties:

CONDITION 3.2. The function  $\gamma : \overline{(a, b)} \rightarrow \mathbb{R}$  is such that

- (i)  $\gamma$  is Borel-measurable and not identically equal to zero,
- (ii)  $\gamma$  is decreasing on  $\overline{(a, b)}$ ,
- (iii)  $E|\gamma(Z)| = \int_a^b |\gamma(t)|p(t)dt < \infty$  and in fact  $E[\gamma(Z)] = \int_a^b \gamma(t)p(t)dt = 0$ .

Henceforth, we will always assume that Conditions 3.1 and 3.2 are satisfied. Note that by Condition 3.2 there exists a point  $x_0 \in (a, b)$  such that

$$(25) \quad \gamma(x) \geq 0 \quad \text{if } a < x < x_0 \quad \text{and} \quad \gamma(x) \leq 0 \quad \text{if } x_0 < x < b,$$

though it might not be unique. For definiteness, we choose

$$(26) \quad x_0 := \sup\{x \in (a, b) : \gamma(x) > 0\}.$$

By item (iii) of Condition 3.2 we can define the function  $I : \overline{(a, b)} \rightarrow \mathbb{R}$  by

$$(27) \quad I(x) := \int_a^x \gamma(t)p(t)dt = - \int_x^b \gamma(t)p(t)dt$$

and by the positivity of  $p$  on  $(a, b)$  we can define the function  $\eta$  on  $(a, b)$  by

$$(28) \quad \eta(x) := \frac{I(x)}{p(x)} = \frac{1}{p(x)} \int_a^x \gamma(t)p(t)dt = - \frac{1}{p(x)} \int_x^b \gamma(t)p(t)dt.$$

The following proposition lists some properties of the function  $I$ .

PROPOSITION 3.3. *Under Conditions 3.1 and 3.2 the function  $I$  has the following properties:*

- (a)  $I$  is locally absolutely continuous on  $\overline{(a, b)}$ .
- (b)  $I$  is nonnegative and  $I(a+) = I(b-) = 0$ .

- (c)  $I$  is increasing on  $\overline{(a, x_0]}$  and decreasing on  $\overline{[x_0, b)}$  and, hence, attains its global maximum at  $x_0$ .

*Proof.* Of course, (a) follows from the fundamental theorem of calculus for Lebesgue integration and the second part of (b) is immediate from item (iii) of Condition 3.2. Finally, (c) and the first part of (b) follow from the second part of (b) and (25).  $\square$

If  $a > -\infty$  and/or  $b < \infty$ , then it is of interest to know under what circumstances it is possible to extend  $\eta$  to a continuous function on  $\overline{(a, b)}$  because we would like to have  $\eta(W)$  make sense, even if  $W$  assumes one of the boundary values  $a$  and  $b$  with positive probability. We will see that in most cases we indeed have  $\eta(a+) = 0$  or  $\eta(b-) = 0$  if  $a > -\infty$  or if  $b < \infty$ , respectively. The following Mills ratio condition is satisfied by most absolutely continuous distributions and will in fact turn out to be equivalent to the asserted boundary behaviour of  $\eta$ . From now on, we will denote by  $F$  the distribution function corresponding to the density  $p$ .

CONDITION 3.4. The density  $p$  of  $Z$  satisfies all the properties from Condition 3.1 and also the following:

- (i) If  $a > -\infty$ , then  $\lim_{x \downarrow a} \frac{F(x)}{p(x)} = 0$ .
- (ii) If  $b < \infty$ , then  $\lim_{x \uparrow b} \frac{1-F(x)}{p(x)} = 0$ .

PROPOSITION 3.5. Assume that Conditions 3.1 and 3.2 hold for  $p$  and  $\gamma$ , respectively. Then, the function  $\eta$  vanishes at the finite end points of the support  $\overline{(a, b)}$  of  $\mathcal{L}(Z)$ , i.e.  $\eta(a+) = 0$  whenever  $a > -\infty$  and  $\eta(b-) = 0$  whenever  $b < \infty$ , if and only if Condition 3.4 is satisfied. Thus, in this case we can extend  $\eta$  to a continuous function on  $\overline{(a, b)}$  vanishing at the finite end points of this interval.

Not every density  $p$  satisfies Condition 3.4 as is clarified by the following example.

EXAMPLE 3.6. Let  $\delta_n \in (0, 1)$ ,  $n \geq 1$ , be such that  $\sum_{n \geq 1} \delta_n = 1$  and define  $x_n := 1 - \sum_{j=1}^{n-1} \delta_j = \sum_{j=n}^{\infty} \delta_j$  and  $I_n := [x_{n+1}, x_n]$ ,  $n \geq 1$ . Furthermore let  $q$  be the unique continuous function, which is linear on each interval  $I_n$  and such that  $q(x_{2n}) = \delta_{2n}^2$  and  $q(x_{2n+1}) = \delta_{2n}$  for  $n \geq 1$  and  $q(1) := \delta_1$ . Define  $p$  to be the probability density which is a constant multiple of  $q$ . Then,  $p$  satisfies Condition 3.1 with  $a = 0$  and  $b = 1$  but Condition 3.4 does not hold: We have  $\lim_{n \rightarrow \infty} x_{2n} = 0$  but

$$\begin{aligned} \frac{F(x_{2n})}{p(x_{2n})} &\geq \frac{F(x_{2n}) - F(x_{2n+1})}{p(x_{2n})} = \frac{1}{p(x_{2n})} \int_{x_{2n+1}}^{x_{2n}} p(t) dt \\ &= \frac{\delta_{2n}(p(x_{2n}) + p(x_{2n+1}))}{2p(x_{2n})} \geq \frac{\delta_{2n}p(x_{2n+1})}{2p(x_{2n})} = \frac{1}{2}. \end{aligned}$$

Note that  $p$  satisfies  $\lim_{x \rightarrow 0} p(x) = 0$ , so that this does not only happen because  $p(0+)$  might not exist.

The counterexample given in Example 3.6 is quite artificial. Indeed, the following proposition lists mild assumptions on the density  $p$  which guarantee that Condition 3.4 is satisfied. In practice, at least one of these assumptions is usually met. In particular, note that by part (f) of Proposition 3.7 the Mills ratio limits from Condition 3.4 at finite boundary points  $a$  or  $b$  are always zero, whenever they exist.

PROPOSITION 3.7. Assume  $a > -\infty$ . In either of the following cases  $\lim_{x \downarrow a} \frac{F(x)}{p(x)} = 0$ .

- (a) The density  $p$  is bounded away from zero in a suitable neighbourhood of  $a$ .
- (b) We have  $p(a+) = 0$  and there is a  $\delta > 0$  such that  $p$  is increasing on  $(a, a + \delta)$ .
- (c) We have  $p(a+) = 0$  and there is a  $\delta > 0$  such that  $p$  is convex on  $(a, a + \delta)$ .
- (d) We have  $p(a+) = 0$  and there is a  $\delta > 0$  such that  $p$  is concave on  $(a, a + \delta)$ .
- (e) The density  $p$  is analytic at  $a$ .
- (f) The limit  $\lim_{x \downarrow a} \frac{F(x)}{p(x)}$  exists.

Of course, similar conditions guarantee that  $\lim_{x \uparrow b} \frac{1-F(x)}{p(x)} = 0$  if  $b < \infty$ .

The proof is given in Section 5. Now, for a given Borel-measurable function  $h$  on  $(a, b)$  such that  $E|h(Z)| < \infty$  consider the Stein equation (18). It is easy to check that the function

$g_h : (a, b) \rightarrow \mathbb{R}$  given by

$$\begin{aligned} g_h(x) &:= \frac{1}{p(x)\eta(x)} \int_a^x (h(t) - E[h(Z)])p(t)dt \\ (29) \quad &= -\frac{1}{p(x)\eta(x)} \int_x^b (h(t) - E[h(Z)])p(t)dt \end{aligned}$$

is a solution to (18) in the sense that  $g_h$  is locally absolutely continuous on  $(a, b)$  and (18) holds for each point  $x \in (a, b)$  where  $g_h$  is in fact differentiable. At all other points  $x \in (a, b)$ , in contrast to the usual convention, we define  $g'_h(x)$  by (18) such that this identity holds true on  $(a, b)$ . The formula for  $g_h$  might as well be found by the method of variation of the constant using the fact that  $\log(p\eta)$  is a primitive function of  $\gamma/\eta$ , which in turn follows from (17). In what follows we will always call the solution  $g_h$  given by (29) the *standard solution* to equation (18). It turns out that this particular solution has the best properties. For instance, if  $g_h$  is bounded then it is the only bounded solution since the solutions of the corresponding homogeneous equation are given by constant multiples of  $I^{-1} = \eta^{-1}p^{-1}$ , which is unbounded by Proposition 3.3 (b). Since we do not exclude cases where the given random variable  $W$  assumes the finite value  $a$  and/or  $b$  with positive probability, we have to make sure that  $g_h$  can be extended to a continuous function on  $\overline{(a, b)}$ . Assume that  $a > -\infty$ . Here, and in what follows we will often write  $\tilde{h}$  for  $h - E[h(Z)]$ . By de l'Hôpital's rule we have

$$(30) \quad \lim_{x \downarrow a} g_h(x) = \lim_{x \downarrow a} \frac{\int_a^x \tilde{h}(t)p(t)dt}{I(x)} = \lim_{x \downarrow a} \frac{\tilde{h}(x)p(x)}{\gamma(x)p(x)} = \lim_{x \downarrow a} \frac{\tilde{h}(x)}{\gamma(x)} = \frac{h(a+) - E[h(Z)]}{\gamma(a+)},$$

if  $h$  has a right limit at  $a$ . Note that  $\gamma$  has a right limit at  $a$  since it is decreasing. Similarly,

$$\lim_{x \uparrow b} g_h(x) = \frac{h(b-) - E[h(Z)]}{\gamma(b-)},$$

if  $h$  has a left limit at  $b < \infty$ .

PROPOSITION 3.8. Assume Conditions 3.1 and 3.2 and let  $h : \overline{(a, b)} \rightarrow \mathbb{R}$  be a Borel-measurable function such that  $E|h(Z)| < \infty$ . Then, the standard solution  $g_h$  to (18) given by (29) has the following properties:

- (a) If  $a > -\infty$  and  $h$  has a right limit at  $a$ , then  $g_h$  can be extended continuously to  $a$  by letting  $g_h(a) := \frac{h(a+) - E[h(Z)]}{\gamma(a+)}$ .
- (b) If  $b < \infty$  and  $h$  has a left limit at  $b$ , then  $g_h$  can be extended continuously to  $b$  by letting  $g_h(b) := \frac{h(b-) - E[h(Z)]}{\gamma(b-)}$ .

The success of Stein's method within applications considerably depends on good bounds on the solutions  $g_h$  and their lower order derivatives, generally uniformly over some given class of test functions  $h$ . The next step will be to prove such bounds. It has to be mentioned that we cannot expect to derive concrete good bounds in full generality, but that sometimes further conditions have to be imposed either on the density  $p$  or on the coefficient  $\gamma$ . Nevertheless, we will derive bounds involving functional expressions which can be simplified, computed or further bounded a posteriori for concrete distributions. So our abstract viewpoint will pay off. Moreover, some of our general bounds will already be explicit.

In what follows, we denote by  $g_h$  the standard solution to Stein's equation (18) on  $\overline{(a, b)}$ , implicitly assuming that  $h$  satisfies the assumptions of Proposition 3.8. Furthermore, for a function  $f$  we denote by  $\|f\|_\infty$  its essential supremum norm on  $\overline{(a, b)}$ . Note that this implies for  $f$  a Lipschitz-continuous function on  $\overline{(a, b)}$  that  $\|f'\|_\infty$  is just its minimum Lipschitz constant. First we give bounds for bounded and measurable test functions  $h$ .

**PROPOSITION 3.9.** *Assume Conditions 3.1 and 3.2 and let  $m$  be the median of  $\mathcal{L}(Z)$ . Then, if  $h : \overline{(a, b)} \rightarrow \mathbb{R}$  is Borel-measurable and bounded we have*

$$(31) \quad \|g_h\|_\infty \leq \frac{\|h - E[h(Z)]\|_\infty}{2I(m)} = \frac{\|h - E[h(Z)]\|_\infty}{2 \int_a^m \gamma(t)p(t)dt}.$$

The proof is deferred to Section 5. The following corollary specializes this result to the case that  $\gamma(x) = -c(x - E[Z])$  and that  $\mathcal{L}(Z)$  is symmetric with respect to its mean  $E[Z]$ , i.e.  $Z - E[Z] \stackrel{D}{=} E[Z] - Z$ . Then, it is also clear that  $m = E[Z]$ .

**COROLLARY 3.10.** *In addition to Conditions 3.1 and 3.2 assume that the distribution  $\mathcal{L}(Z)$  is symmetric with respect to  $m = E[Z]$  and that  $\gamma(x) = -c(x - E[Z])$  for some positive constant  $c$ . Then, for each bounded and Borel-measurable test function  $h : \overline{(a, b)} \rightarrow \mathbb{R}$  we have*

$$(32) \quad \|g_h\|_\infty \leq \frac{\|h - E[h(Z)]\|_\infty}{cE[|Z - m|]}.$$

*Proof.* In this case we clearly have  $I(m) = \frac{c}{2}E[|Z - m|]$  which implies the result by Proposition 3.9. □

In the case that  $Z \sim N(0, 1)$  and  $c = 1$  this result specializes to the well known bound  $\|g_h\|_\infty \leq \sqrt{\frac{\pi}{2}}\|h - E[h(Z)]\|_\infty$  (see [7] or [8], e.g.).

**REMARK 3.11.** (a) In the statement of Proposition 3.9 it might surprise that there is no bound mentioned for  $\|g'_h\|_\infty$ . This is because, in general, a bound of the form  $\|g'_h\|_\infty \leq C\|h - E[h(Z)]\|_\infty$  with a finite constant  $C$  does not exist in this

setup. For instance, for  $z > 0$  and  $Z$  having the exponential distribution with mean one, consider the Stein equation

$$(33) \quad xg'(x) + (1-x)g(x) = 1_{[0,z]}(x) - P(Z \leq z).$$

Identity (3.3) from [4] shows that for  $x > z$  the solution  $g_z$  to (33) satisfies

$$g'_z(x) = \frac{e^{-z} - 1}{x^2}.$$

Hence, we have that

$$\sup_{x>z} |g'_z(x)| = \frac{1 - e^{-z}}{z^2} \xrightarrow{z \downarrow 0} \infty,$$

proving that such a constant  $C$  in general cannot exist. Note also that this is contrary to the density approach, where one usually has such a bound (see [5] or [7]).

- (b) The Kolmogorov distance between a given random variable  $W$  and  $Z$  is induced by the class of test functions  $h_z := 1_{(-\infty, z]}$ , where  $z \in (a, b)$ . In this situation it is easy to verify that the standard solution  $g_z := g_{h_z}$  to (18) is given by

$$g_z(x) = \begin{cases} \frac{F(x)(1-F(z))}{I(x)}, & a < x \leq z \\ \frac{F(z)(1-F(x))}{I(x)}, & z < x < b \end{cases} \quad \text{and} \quad \|g_z\|_\infty = \frac{F(z)(1-F(z))}{I(z)} =: S(z).$$

By using de l'Hôpital's rule it is not hard to check that always  $\sup_{z \in (a,b)} S(z) < \infty$ . Furthermore,  $g_z$  is Lipschitz-continuous and on  $(a, b) \setminus \{z\}$  it is infinitely often continuously differentiable with

$$g'_z(x) = \begin{cases} \frac{(1-F(z))p(x)H(x)}{I(x)^2}, & a < x < z \\ \frac{-F(z)p(x)G(x)}{I(x)^2}, & z < x < b, \end{cases}$$

where the functions  $H$  and  $G$  are defined in Proposition 3.13. From the negative example of (a) we already know that, in general, there is no finite constant  $C$  such that

$$\|g'_z\|_\infty \leq C, \quad a < z < b.$$

Nevertheless, even in such a situation, one may use the uniform bound on  $S$  and a  $z$ -dependent bound on  $\|g'_z\|_\infty$  as well as particular properties of  $W$  to prove accurate bounds on the Kolmogorov distance. This was done in [4] for the exponential distribution. Incidentally, in the case of the Beta distribution, the function  $S$  will be bounded for a different purpose in the proof of Proposition 4.2.

Proposition 3.9 is already sufficient to prove that the operator  $L$  given by (15) characterizes the distribution of  $Z$ . The proof is given in Section 5.

**PROPOSITION 3.12.** *A random variable  $X$  with values in  $\overline{(a, b)}$  has the same distribution as  $Z$  if and only if for each continuous function  $f$  on  $\overline{(a, b)}$ , which is locally absolutely continuous on  $(a, b)$  and which satisfies*

$$E|\eta(Z)f'(Z)| = \int_a^b |f'(x)|I(x)dx < \infty \text{ we have}$$

$$E[\eta(X)f'(X)] = -E[\gamma(X)f(X)].$$

*In particular, in this case both expected values exist.*

Next, we will turn to Lipschitz continuous test functions  $h$ . In contrast to bounded measurable test functions, there we will also be able to prove useful bounds for  $g'_h$ . In order that  $E[h(Z)]$  exists for Lipschitz continuous test functions  $h$  we need to assume that  $E|Z| < \infty$ . The following two result, which are also proved in Section 5, include optimal bounds for both,  $g_h$  and  $g'_h$ , when  $h$  is Lipschitz.

**PROPOSITION 3.13.** *Assume that  $E|Z| < \infty$  and Conditions 3.1 and 3.2 hold. Then, we have for any Lipschitz continuous test function  $h : \overline{(a, b)} \rightarrow \mathbb{R}$  and any  $x \in \overline{(a, b)}$ :*

$$\begin{aligned} \text{(a)} \quad |g_h(x)| &\leq \|h'\|_\infty \frac{F(x)E[Z] - \int_a^x yp(y)dy}{I(x)} = \|h'\|_\infty \frac{\int_a^x (E[Z] - y)p(t)dt}{I(x)}; \\ \text{(b)} \quad |g'_h(x)| &\leq \|h'\|_\infty \frac{\int_a^x F(s)dsG(x) + \int_x^b (1 - F(s))dsH(x)}{p(x)\eta(x)^2}. \end{aligned}$$

Here, for  $x \in \overline{(a, b)}$ , the positive functions  $H(x)$  and  $G(x)$  are defined by

$$H(x) := I(x) - \gamma(x)F(x) = p(x)\eta(x) - \gamma(x)F(x) \text{ and } G(x) := H(x) + \gamma(x).$$

Moreover, these bounds are optimal among all bounds involving the factor  $\|h'\|_\infty$ .

**REMARK 3.14.** (a) If  $a > -\infty$  and  $b < \infty$ , then it follows by an application of de l'Hôpital's rule that the function  $S(x) := \frac{\int_a^x (E[Z] - y)p(t)dt}{I(x)}$  is bounded on  $(a, b)$ . Indeed, if  $a > -\infty$ , for instance, we have that

$$0 \leq \lim_{x \downarrow a} \frac{\int_a^x (E[Z] - y)p(t)dt}{I(x)} = \lim_{x \downarrow a} \frac{E[Z] - x}{\gamma(x)} = \frac{E[Z] - a}{\gamma(a)} < \infty.$$

However, in general  $S(x)$  is unbounded, if  $|\gamma(x)|$  does not grow at least linearly with  $x$ . For instance, if  $Z \sim N(0, 1)$  and  $\gamma(t) = -\text{sign}(t)$ , then we have for positive  $x$  that

$$S(x) = \frac{\varphi(x)}{1 - \Phi(x)} \sim x$$

by the Gaussian Mills ratio inequality.

(b) The bound for  $|g_h(x)|$  in part (a) of Proposition 3.13 can be written as

$$|g_h(x)| \leq \|h'\|_\infty \frac{\tau(x)}{\eta(x)},$$

where  $\tau$  is the so-called *Stein factor* or *Stein kernel* of  $Z$  given by

$$\tau(x) = \frac{1}{p(x)} \int_a^x (E[Z] - t)p(t)dt,$$

i.e.  $\tau$  is the function  $\eta$  which belongs to the choice  $\gamma(x) = E[Z] - x$ . The Stein kernel  $\tau$  appeared first in Lecture 6 of [32] and it has turned out to be a fundamental object in Stein's method for one-dimensional absolutely continuous distributions (see, e.g. [23], [24] and [21]).

**COROLLARY 3.15.** *Assume that  $E|Z| < \infty$ , Condition 3.1 holds and that  $\gamma(x) = c(E[Z] - x)$  for some  $c > 0$ . Then we have for any Lipschitz continuous test function  $h : \overline{(a, b)} \rightarrow \mathbb{R}$  and each  $x \in (a, b)$ :*

$$\text{(a)} \quad \|g_h\|_\infty \leq \frac{\|h'\|_\infty}{c};$$



$$(b) \quad |g'_h(x)| \leq \frac{2\|h'\|_\infty}{c} \frac{H(x)G(x)}{\eta(x)^2 p(x)} = 2c\|h'\|_\infty \frac{\int_a^x F(s)ds \int_x^b (1-F(t))dt}{\eta(x)^2 p(x)}.$$

REMARK 3.16. (i) In the case of the normal distribution (via its classical Stein equation) the bound given in Corollary 3.15 (a) reduces to  $\|g_h\|_\infty \leq \|h'\|_\infty$ . Formally, this bound is a special instance of a general bound given in Lemma 3.1 of [19] for the multivariate standard normal distribution (see also Lemma 2.6 in [7]). However, this lemma is stated under the additional assumption that  $h$  has three bounded derivatives, which is stronger than being Lipschitz-continuous. Yet, as has been pointed out to me by the referee, one can use the generator representation of the solution to the Stein equation to obtain the same bound as in Corollary 3.15 (a) for once differentiable test functions  $h$  with bounded first derivative by applying the well-known consequences of the dominated convergence theorem on differentiating under the integral sign. Then, using smoothing techniques, this result could be extended to the class of Lipschitz-continuous test functions, yielding an alternative proof of this bound. Nevertheless, in the context of Stein's method for the univariate normal distribution, the best bound mentioned on  $g_h$  for a Lipschitz test function  $h$  is  $\|g_h\|_\infty \leq 2\|h'\|_\infty$  (see, e.g. [7] or [8]). Hence, we believe that Corollary 3.15 (a) is the first result that rigorously proves the aforementioned bound, although, as described above, it can also be proved by means of existing techniques from the generator framework.

- (ii) For concrete distributions the ratio appearing in the bounds for  $g'_h(x)$  may be bounded uniformly in  $x$  by some constant which can sometimes also be computed explicitly. For instance, this is performed for the Beta distribution in Section 4. Furthermore, for the situation of Corollary 3.15, in [12] the authors give mild conditions for the existence of a finite constant  $k$  such that  $\|g'_h\|_\infty \leq k\|h'\|_\infty$  for any Lipschitz-continuous  $h$ . In practice, these conditions are usually met. However, there is no hope of estimating the constant  $k$  by their method of proof. Thus, for concrete distributions and explicit constants it might therefore be useful to work with our bounds from Corollary 3.15 (b) or from Proposition 3.13.
- (iii) For the normal distribution and also for the larger class of distributions discussed in [13], one also has a bound of the form  $\|g''_h\|_\infty \leq C\|h'\|_\infty$  for some finite constant  $C$  holding for each Lipschitz function  $h$ . As was shown by a universal counterexample in [12], if  $\gamma(x) = c(E[Z] - x)$  such a bound cannot be expected unless  $a = -\infty$  and  $b = \infty$ . If either  $a > -\infty$  or  $b < \infty$  one will have to assume that  $h'$  is also Lipschitz, for example by demanding that  $h$  has two bounded derivatives, in order to obtain a finite bound on  $\|g''_h\|_\infty$ . Within the density approach, however, there are many examples of distributions, whose support is strictly included in  $\mathbb{R}$  but for which such bounds are available (see, e.g., chapter 13 of [7]).

Now, we show how we can use the above results and the density formula (24) to give bounds on higher order derivatives of  $g_h$ , if  $h$  itself is smooth enough. First note that the constant  $K$  from (24) is given by

$$(34) \quad K = \eta(x_0)p(x_0) = \int_a^{x_0} |\gamma(t)|p(t)dt = \int_{x_0}^b |\gamma(t)|p(t)dt = \frac{E|\gamma(Z)|}{2}$$

and, hence, we have the explicit formula

$$(35) \quad p(x) = \frac{E|\gamma(Z)|}{2\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right).$$

Formula (35) is a more general version of formula (3.14) in [24] and is also derived in [20]. Now, if the coefficient  $\gamma$  is also absolutely continuous, by differentiating Stein's equation (18), we obtain for  $h$  Lipschitz

$$(36) \quad \eta(x)g_h''(x) + g_h'(x)(\eta'(x) + \gamma(x)) = h'(x) - \gamma'(x)g_h(x) =: h_2(x).$$

This means, that the function  $\tilde{g} := g_h'$  is a solution of the Stein equation corresponding to the test function  $h_2$  for the distribution of  $\tilde{Z}$  which satisfies the Stein identity

$$E\left[\eta(\tilde{Z})f'(\tilde{Z}) + (\eta'(\tilde{Z}) + \gamma(\tilde{Z}))f(\tilde{Z})\right] = 0.$$

From (35) we know that a density  $\tilde{p}$  of  $\tilde{Z}$  is given by

$$(37) \quad \tilde{p}(x) = \frac{\tilde{K}}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\eta'(t) + \gamma(t)}{\eta(t)} dt\right) = K \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) = C\eta(x)p(x),$$

where  $\tilde{K}, K, C > 0$  are suitable normalizing constants. Thus, if we have bounds for the first derivative of the Stein solutions for the distribution of  $\tilde{Z}$  and for Lipschitz functions  $h$ , then from this observation we obtain bounds on  $g_h''$  for  $h$  such that  $h_2$  is Lipschitz. Note that if  $\gamma(x) = c(E[Z] - x)$ , this essentially means that  $h'$  must be Lipschitz as well. Of course, in order to apply this procedure, one has to make sure that  $E[h_2(\tilde{Z})] = 0$  and that  $g_h'$  is the standard solution to the Stein equation for  $\mathcal{L}(\tilde{Z})$  and the test function  $h_2$ . Remarkably, under mild conditions this turns out to always be the case.

**PROPOSITION 3.17.** *Assume that Conditions 3.1 and 3.2 hold,  $E|Z| < \infty$ ,  $E|Z\gamma(Z)| < \infty$  and that  $\gamma$  is locally absolutely continuous on  $(a, b)$ . Furthermore, let  $h$  be a Lipschitz-continuous function. Then,  $E[h_2(\tilde{Z})]$  exists and equals 0. Furthermore, if either the derivative  $g_h'$  of  $g_h$  is bounded,  $a > -\infty$  or  $b < \infty$ , then  $g_h'$  is the standard solution to the Stein equation*

$$(38) \quad \eta(x)f'(x) + (\eta'(x) + \gamma(x))f(x) = h_2(x)$$

corresponding to the distribution of  $\tilde{Z}$  and the test function  $h_2 = h' - \gamma'g_h$ .

The proof is given in Section 5.

**REMARK 3.18.** If  $Z \sim \text{Beta}(a, b)$ , then (37) implies that  $\tilde{Z} \sim \text{Beta}(a+1, b+1)$ . This will be used in Section 4 to provide bounds on higher order derivatives of  $g_h$  in the case of the Beta distribution. If, on the other hand,  $Z \sim \text{Exp}(\alpha)$  has an exponential distribution with mean  $\alpha^{-1}$ , then  $\tilde{Z} \sim \text{Gamma}(2, \alpha)$  has an Erlang distribution. Using this fact, Proposition 3.17 and the general bounds from Corollary 3.15 applied for both the exponential and the  $\text{Gamma}(2, \alpha)$  distribution, one can, with some work, derive the following bounds on the standard solution  $g_h$  to the Stein equation

$$xg'(x) + (1 - \alpha x)g(x) = h(x) - Eh(Z_\alpha)$$

for the distribution  $\text{Exp}(\alpha)$ , if  $h$  is continuously differentiable on  $[0, \infty)$  and both  $h$  and  $h'$  are Lipschitz:

$$\|g_h\|_\infty \leq \frac{1}{\alpha}\|h'\|_\infty, \quad \|g'_h\|_\infty \leq \|h'\|_\infty \quad \text{and} \quad \|g''_h\|_\infty \leq \frac{2\alpha}{3}\|h'\|_\infty + \frac{2}{3}\|h''\|_\infty$$

These bounds are better than those derived in [14] and, additionally, since we do not have to assume that  $h'(0) = 0$  for the bound on  $\|g''_h\|_\infty$  to be valid, one term in the bounds of Theorems 1.1 and 1.2 from [14] would drop off, if instead our bounds were used.

Next, we introduce the approach of exchangeable pairs satisfying the regression properties (14) and (23) in our general framework. As was observed in [29] for the normal distribution, in case of univariate distributional approximations, one does not need the full strength of exchangeability, but equality in distribution of the random variables  $W$  and  $W'$  is sufficient. This may allow for a greater choice of admissible couplings in several situations, or at least, relaxes the verification of asserted properties. Thus, let  $W, W'$  be real-valued random variables defined on the same probability space such that  $W \stackrel{\mathcal{D}}{=} W'$ . We will assume, that the random variables  $W$  and  $W'$  only have values in an interval  $(a, b) \subseteq J \subseteq \overline{(a, b)}$  where both functions  $\eta$  and  $\gamma$  are defined (recall that it might be the case that  $\eta$  can only be defined on  $(a, b)$ ). However, from Proposition 3.5 we know that we can let  $J = \overline{(a, b)}$  if Condition 3.4 holds.

**PROPOSITION 3.19.** *Assume that Conditions 3.1 and 3.2 hold and that  $W$  is square integrable with  $E|\gamma(W)| < \infty$ . Furthermore, for some constant  $\lambda > 0$ , assume that the general regression property (14) and also the second moment condition (23) are satisfied by the pair  $(W, W')$ . Let  $f : J \rightarrow \mathbb{R}$  be a bounded, continuously differentiable function, which is Lipschitz-continuous and has a Lipschitz-continuous derivative  $f'$ . Then,*

$$(39) \quad \left| E[\eta(W)f'(W) + \gamma(W)f(W)] \right| \leq \frac{\|f''\|_\infty}{6\lambda} E[|W' - W|^3] + \|f\|_\infty E|R| + \|f'\|_\infty E|S|,$$

where  $\|f''\|_\infty$  denotes the minimum Lipschitz constant of  $f'$ .

The proof is given in Section 5.

**REMARK 3.20.** (i) The bound (39) can only be small, if  $S$  and  $R$  are of negligible order.

(ii) The proof shows, that Proposition 3.19 can easily be generalized to the situation, where there is a sub- $\sigma$ -algebra  $\mathcal{F}$  with  $\sigma(W) \subseteq \mathcal{F}$  and the more general regression properties

$$(40) \quad \frac{1}{\lambda} E[W' - W | \mathcal{F}] = \gamma(W) + R \quad \text{and} \quad \frac{1}{2\lambda} E[(W' - W)^2 | \mathcal{F}] = \eta(W) + S$$

hold for some  $\mathcal{F}$ -measurable remainder terms  $R$  and  $S$ .

(iii) If  $\mathcal{H}$  is some class of test functions, such that there are finite, positive constants  $c_0$ ,  $c_1$  and  $c_2$  with  $\|g_h\|_\infty \leq c_0$ ,  $\|g'_h\|_\infty \leq c_1$  and  $\|g''_h\|_\infty \leq c_2$  for each  $h \in \mathcal{H}$ , then (39) immediately yields a bound on the distance

$$d_{\mathcal{H}}(\mathcal{L}(Z), \mathcal{L}(W)) = \sup_{h \in \mathcal{H}} \left| E[h(W)] - E[h(Z)] \right|.$$

Finally, in our general framework, we readdress the last issue discussed in Section 2. Namely, we suppose that we are given two functions  $\gamma$  and  $\eta$ , such that for some  $-\infty \leq a < b \leq \infty$  the function  $\gamma$  is defined on  $\overline{(a, b)}$ ,  $\eta$  is defined at least on  $(a, b)$  and the following properties hold.

- CONDITION 3.21. (a) The function  $\gamma$  is decreasing and such that  $0 < \gamma(a+) \leq \infty$  and  $-\infty \leq \gamma(b-) < 0$ . Again, we define  $x_0 \in (a, b)$  by  $x_0 := \sup\{x \in (a, b) : \gamma(x) > 0\}$ .  
 (b) The function  $\eta$  is positive and locally absolutely continuous on  $(a, b)$ .  
 (c) The function  $\gamma/\eta$  is locally integrable on  $(a, b)$  and, if we define

$$Q(x) := \int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt, \quad x \in (a, b),$$

then we have  $Q(a+) = Q(b-) = -\infty$ .

Note that by definition we have  $Q(x) \leq 0$  for all  $x \in (a, b)$ , if Condition 3.21 is satisfied. Now, we define the density  $p$  by relation (24) with  $K$  being a suitable normalizing constant. The existence of  $K$  follows from the fact that, by Condition 3.21, for each  $c \in (a, x_0)$  there is a finite constant  $L > 0$  such that  $L\gamma(x) \geq 1$  for each  $x \in (a, c)$ . Thus,

$$\begin{aligned} \int_a^c \frac{1}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) dx &\leq L \int_a^c \frac{\gamma(x)}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) dx \\ &= L \int_a^c Q'(x) \exp(Q(x)) dx = L \int_{-\infty}^{Q(c)} e^u du \\ (41) \quad &= L e^{Q(c)} < \infty. \end{aligned}$$

A similar calculation shows that also

$$\int_d^b \frac{1}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) dx < \infty$$

for each  $d \in (x_0, b)$ . Hence,  $p$  can be suitably normalized. Now, let  $Z$  be a random variable with probability density function  $p$ . The next result is a generalization of Lemma 3, Lecture 6 in [32].

**THEOREM 3.22.** *If Condition 3.21 is satisfied, then the density  $p$  defined by (24) is such that*

$$E[\gamma(Z)] = \int_a^b \gamma(t)p(t)dt = 0 \quad \text{and} \quad \eta(x) = \frac{1}{p(x)} \int_a^x \gamma(t)p(t)dt, \quad a < x < b.$$

*In particular, the theory developed in this section can be applied in this framework.*

*Proof.* Similarly to (41) we obtain

$$\begin{aligned} \int_a^{x_0} \gamma(x)p(x)dx &= K \int_a^{x_0} \frac{\gamma(x)}{\eta(x)} \exp\left(\int_{x_0}^x \frac{\gamma(t)}{\eta(t)} dt\right) dx \\ &= K e^{Q(x_0)} = K \end{aligned}$$

and

$$\int_{x_0}^b \gamma(x)p(x)dx = -K e^{Q(x_0)} = -K.$$

Thus,  $E[\gamma(Z)] = 0$ . The second claim follows from

$$\begin{aligned} \frac{1}{p(x)} \int_a^x \gamma(t)p(t)dt &= \frac{1}{p(x)} K \int_a^x \frac{\gamma(s)}{\eta(s)} \exp\left(\int_{x_0}^s \frac{\gamma(t)}{\eta(t)} dt\right) ds \\ &= \frac{1}{p(x)} K e^{\mathcal{Q}(x)} = \eta(x), \end{aligned}$$

since

$$p(x) = \frac{K}{\eta(x)} e^{\mathcal{Q}(x)}.$$

□

#### 4. STEIN'S METHOD FOR THE BETA DISTRIBUTION

In this section we specialize the theory from Section 3 to the family  $Beta(a, b)$ ,  $a, b > 0$ , of Beta distributions as defined in Section 2. Let us fix  $a, b > 0$  and from now on assume that  $Z \sim Beta(a, b)$ . Motivated by the Pólya urn example, the above constructed exchangeable pair  $(W, W')$  and by Proposition 2.2 we define the function  $\gamma := \gamma_{a,b}$  as in Proposition 2.2 and observe that

$$E[\gamma(Z)] = 0 \quad \text{since} \quad E[Z] = \frac{a}{a+b}.$$

It is thus easy to see that  $\gamma$  satisfies all assumptions of Condition 3.2 and also that the Beta density  $p := p_{a,b}$  given by (4) satisfies Conditions 3.1 and 3.4, the latter either directly or by Proposition 3.7. We claim that the function  $\eta$  defined by (21) is given by

$$(42) \quad \eta(x) = x(1-x), \quad x \in [0, 1].$$

This is equivalent to proving that

$$(43) \quad p(x)x(1-x) = \int_0^x (a - (a+b)t)p(t)dt, \quad 0 < x < 1,$$

which easily follows from differentiating both sides of (43) and using (10). Thus, from Proposition 3.12 we immediately obtain the following Stein characterization for the Beta distribution. This result substantially extends Theorem 1 in [30] in the case of the Beta distribution, which is weaker as it only characterizes the Beta distribution among the class of absolutely continuous distributions with finite second moment.

**PROPOSITION 4.1.** *A random variable  $X$  with values in  $[0, 1]$  has the distribution  $Beta(a, b)$  if and only if for each continuous function  $f$  on  $[0, 1]$ , which is locally absolutely continuous on  $(0, 1)$  such that  $E|Z(1-Z)f'(Z)| < \infty$ , we have*

$$E\left[X(1-X)f'(X)\right] = (a+b)E\left[\left(X - \frac{a}{a+b}\right)f(X)\right].$$

For the Beta distribution and a measurable function  $h$  with  $E|h(Z)| < \infty$ , the Stein equation (18) is given by

$$(44) \quad x(1-x)g'(x) + (a+b)\left(\frac{a}{a+b} - x\right)g(x) = h(x) - E[h(Z)], \quad x \in [0, 1]$$

and the standard solution (29) has the form

(45)

$$g_h(x) = \frac{1}{x(1-x)p(x)} \int_0^x \tilde{h}(t)p(t)dt = \frac{-1}{x(1-x)p(x)} \int_x^1 \tilde{h}(t)p(t)dt, \quad 0 < x < 1$$

where, again,  $\tilde{h}(t) = h(t) - E[h(Z)]$  and

$$(46) \quad g_h(0) = \frac{h(0+) - E[h(Z)]}{a} \quad \text{and} \quad g_h(1) = \frac{h(1-) - E[h(Z)]}{-b},$$

if  $h$  has a right limit at 0 and a left limit at 1 by Proposition 3.8. We mention that the same Stein equation (44) has already been considered in [30], [18] and in [10].

For  $a, b > 0$  define the constant

$$(47) \quad C(a, b) = 2(a+b) \begin{cases} B(a, b), & a \leq 1, b \leq 1 \\ a^{-1}, & a \leq 1, b > 1 \\ b^{-1}, & a > 1, b \leq 1 \\ a^{-1}b^{-1}B(a, b)^{-1}, & a > 1, b > 1 \end{cases} \quad \text{if } a \neq b \quad \text{and}$$

$$(48) \quad C(a, a) = \begin{cases} 4, & 0 < a < 1 \\ \frac{2a\sqrt{\pi}\Gamma(a)}{\Gamma(a+1/2)}, & a \geq 1. \end{cases}$$

From Proposition 3.9 and Corollary 3.15 we can derive the following bounds for the solution (45) to (44). The proof is given in Section 5.

**PROPOSITION 4.2.** *Let  $h : [0, 1] \rightarrow \mathbb{R}$  be Borel-measurable with  $E|h(Z)| < \infty$ .*

- (a) *If  $h$  is bounded, then  $\|g_h\|_\infty \leq \frac{\|h - E[h(Z)]\|_\infty}{2m(1-m)p(m)}$ , where  $m$  is the median of  $\text{Beta}(a, b)$ .*
- (b) *If  $h$  is Lipschitz, then  $\|g_h\|_\infty \leq \frac{\|h'\|_\infty}{a+b}$  and  $\|g'_h\|_\infty \leq C(a, b)\|h'\|_\infty$ , where  $C(a, b)$  is given by (47) and (48).*
- (c) *If  $h$  is continuously differentiable with Lipschitz derivative  $h'$ , then  $g'_h$  is Lipschitz and  $\|g''_h\|_\infty \leq C(a+1, b+1)\|h''\|_\infty + (a+b)C(a+1, b+1)C(a, b)\|h'\|_\infty$ .*
- (d) *More generally, if  $m \geq 1$  is an integer and  $h$  is at least  $(m-1)$ -times differentiable such that  $h^{(j)}$  is Lipschitz-continuous for  $j = 0, \dots, m-1$ , then  $\|g_h^{(m-1)}\|_\infty$  is Lipschitz and*

$$\begin{aligned} \|g_h^{(m)}\|_\infty &\leq C(a+m-1, b+m-1) \\ &\cdot \sum_{j=1}^m \left( \prod_{l=j}^{m-1} (l(a+b+l-1)C(a+l-1, b+l-1)) \right) \|h^{(j)}\|_\infty, \end{aligned}$$

where we define an empty product to be equal to 1.

**REMARK 4.3.** (i) It is worthwhile to compare our bound for  $\|g'_h\|_\infty$  from Proposition 4.2 (b) to the bound  $\|g'_h\|_\infty \leq (b_0 + b_1)\|h'\|_\infty$  given in [18]. One can show that if  $a = b$ , then our bound is uniformly better than theirs. However, if  $a \neq b$ , then there are regions for  $(a, b)$  where our constant  $C(a, b)$  is smaller and other ones, where their  $b_0 + b_1$  is smaller. For instance, if  $0 < a, b \leq 1$ , then, again,  $C(a, b) \leq b_0 + b_1$ . But, if  $1 < b < 2$  is fixed and  $a$  tends to zero, then  $C(a, b)$

goes to infinity while their  $b_0 + b_1$  tends to 12. In any case, neither our bound nor the bound from [18] seem to be optimal for  $\|g'_h\|_\infty$ .

(ii) From Corollary 3.15 (b) we know that for Lipschitz  $h$  and  $x \in (0, 1)$

$$|g'_h(x)| \leq \frac{2\|h'\|_\infty}{a+b} \frac{H(x)G(x)}{x^2(1-x)^2p(x)} =: \|h'\|_\infty B(x).$$

By an application of de l'Hôpital's rule, one can show that

$$B(0+) = \frac{2}{a+1} \quad \text{and} \quad B(1-) = \frac{2}{b+1}.$$

We conjecture that if  $\min(a, b) < 1$ , then

$$\|B\|_\infty = \frac{2}{\min(a, b) + 1},$$

i.e. that  $B$  assumes its maximum value at the boundary of  $(0, 1)$ . However, if  $\min(a, b) > 1$ , then we believe that there is always an  $x_1 \in (0, 1)$  such that

$$\frac{2\|h'\|_\infty}{a+b} \frac{H(x_1)G(x_1)}{x_1^2(1-x_1)^2p(x_1)} > \frac{2\|h'\|_\infty}{\min(a, b) + 1}.$$

(iii) If  $a = b$ , then the median of  $Beta(a, a)$  equals  $1/2$  and the bound in (a) has the explicit form  $\|g_h\|_\infty \leq B(a, a)2^{a+b-1}\|h - E[h(Z)]\|_\infty$ . Unfortunately, for  $a \neq b$  there is no closed form expression for the median of  $Beta(a, b)$ . In such a case one could use known inequalities about the median  $m$  in order to get bounds on  $\|g_h\|_\infty$ . Since one would have to distinguish several cases according to the values of  $a$  and  $b$  and, hence, to the shape of the density  $p$ , we omit the details, here.

From Proposition 3.19, Remark 3.20 (ii) and the bounds from Proposition 4.2 we obtain the following plug-in result, which bounds a certain distance to the Beta distribution by terms related to a given exchangeable pair.

**THEOREM 4.4.** *Let  $W$  and  $W'$  be identically distributed random variables on a common probability space  $(\Omega, \mathcal{A}, P)$  and let  $\mathcal{F} \subseteq \mathcal{A}$  be a sub- $\sigma$ -algebra of  $\mathcal{A}$  such that  $\sigma(W) \subseteq \mathcal{F}$  and*

$$\begin{aligned} \frac{1}{\lambda} E[W' - W \mid \mathcal{F}] &= (a+b) \left( \frac{a}{a+b} - W \right) + R \quad \text{and} \\ \frac{1}{2\lambda} E[(W' - W)^2 \mid \mathcal{F}] &= W(1-W) + S \end{aligned}$$

*hold for a constant  $\lambda > 0$  and for  $\mathcal{F}$ -measurable remainder terms  $R$  and  $S$ . Then, for each continuously differentiable function  $h : [0, 1] \rightarrow \mathbb{R}$  with a Lipschitz derivative  $h'$  it holds that*

$$\begin{aligned} |E[h(W)] - E[h(Z)]| &\leq \|h'\|_\infty \left( \frac{1}{a+b} E|R| + C(a, b) E|S| \right) \\ &+ \left( \frac{C(a+1, b+1)\|h''\|_\infty + (a+b)C(a+1, b+1)C(a, b)\|h'\|_\infty}{6\lambda} \right) E|W' - W|^3, \end{aligned}$$

where the constants  $C(\cdot, \cdot)$  are defined by (47) and (48).

Now we are in a position to prove Theorem 2.1.

*Proof of Theorem 2.1.* The claim immediately follows from Theorem 4.4, Propositions 2.2, 2.3 and the fact that in this case

$$R = 0, \quad S = \frac{b-a}{2n}W + \frac{a}{2n} \geq 0 \quad \text{and} \quad |W' - W| \leq \frac{1}{n}.$$

□

## 5. PROOFS

*Proof of Proposition 3.5.* Suppose, that  $a > -\infty$  and choose  $y \in (a, x_0)$ . Then  $\gamma(y) > 0$  and, by the nonnegativity of  $I$  and the monotonicity of  $\gamma$ , for  $a < x < y$  we have

$$\begin{aligned} 0 &\leq \gamma(y)F(x) = \gamma(y) \int_a^x p(t)dt \leq I(x) = \int_a^x \gamma(t)p(t)dt \\ (49) \quad &\leq \gamma(a) \int_a^x p(t)dt = \gamma(a)F(x). \end{aligned}$$

Hence, if Condition 3.4 holds, we have

$$0 \leq \liminf_{x \downarrow a} \eta(x) \leq \limsup_{x \downarrow a} \eta(x) = \limsup_{x \downarrow a} \frac{I(x)}{p(x)} \leq \gamma(a) \lim_{x \downarrow a} \frac{F(x)}{p(x)} = 0,$$

so that  $\lim_{x \downarrow a} \eta(x) = 0$ . Conversely, if  $\eta(a+) = 0$ , then, again by (49),

$$0 \leq \liminf_{x \downarrow a} \frac{F(x)}{p(x)} \leq \limsup_{x \downarrow a} \frac{F(x)}{p(x)} \leq \frac{1}{\gamma(y)} \limsup_{x \downarrow a} \frac{I(x)}{p(x)} = \frac{1}{\gamma(y)} \limsup_{x \downarrow a} \eta(x) = 0.$$

The calculation for finite  $b$  is similar by using the representation  $I(x) = -\int_x^b \gamma(t)p(t)dt$  and is therefore omitted.

□

*Proof of Proposition 3.7.* That item (a) is sufficient is clear. If (b) holds, then the claim follows from the inequality

$$F(x) = \int_a^x p(t)dt \leq p(x)(x-a),$$

valid for  $x \in (a, a + \delta)$ . Under Condition (c) we obtain a continuous and convex function on  $[a, a + \delta)$  by letting  $p(a) := 0$ . Now, let  $a < x < y < a + \delta$ . Then, there exists a  $\lambda \in (0, 1)$  with  $x = \lambda a + (1 - \lambda)y$  and by convexity we have:

$$\begin{aligned} p(y) - p(x) &= p(y) - p(\lambda a + (1 - \lambda)y) \geq p(y) - \lambda p(a) - (1 - \lambda)p(y) \\ &= \lambda p(y) > 0. \end{aligned}$$

Thus, the assumptions of (b) are satisfied. If (d) holds, then again letting  $p(a) := 0$  we obtain a continuous and concave function on  $[a, a + \delta)$ . Thus, there exists a decreasing function  $f$  on  $[a, a + \delta)$  such that

$$p(x) = \int_a^x f(t)dt, \quad a \leq x < a + \delta.$$



If there was a sequence  $(x_n)_{n \geq 1}$  in  $[a, a + \delta)$  such that  $x_n \downarrow a$  and  $f(x_n) \leq 0$  for each  $n \geq 1$ , then for each  $x \in (a, a + \delta)$  and large enough  $n$  we would have

$$p(x) = p(x_n) + \int_{x_n}^x f(t)dt \leq p(x_n) + (x - x_n)f(x_n) \leq p(x_n) \xrightarrow{n \rightarrow \infty} 0,$$

which would contradict Condition 3.1. Thus, there is an  $\varepsilon < \delta$  such that  $f(x) \geq 0$  for all  $x \in (a, a + \varepsilon)$ . Hence,  $p$  is increasing on  $(a, a + \varepsilon)$  and (b) is satisfied. If (e) holds, then there is an  $r > 0$  and a real sequence  $(c_k)_{k \geq 0}$  such that  $p(x) = \sum_{k=0}^{\infty} c_k(x-a)^k$  for all  $x \in (a, a+r)$  and the function  $f : (a-r, a+r) \rightarrow \mathbb{R}$  with  $f(x) := \sum_{k=0}^{\infty} c_k(x-a)^k$  is well-defined. Let  $n_0 := \min\{k \geq 0 : c_k \neq 0\}$ . Then  $n_0 < \infty$  since  $p$  is positive on  $(a, b)$ . If  $n_0 = 0$  and hence  $f(a) = c_0 = \lim_{x \downarrow a} p(x) \neq 0$ , then there is nothing to show. Otherwise, we have

$$p(x) = (x-a)^{n_0} \sum_{k=n_0}^{\infty} c_k(x-a)^{k-n_0} \quad \text{and} \quad p'(x) = (x-a)^{n_0-1} \sum_{k=n_0}^{\infty} k c_k(x-a)^{k-n_0}$$

for all  $x \in (a, a+r)$  and hence, by de l'Hôpital's rule,

$$\begin{aligned} \lim_{x \downarrow a} \frac{F(x)}{p(x)} &= \lim_{x \downarrow a} \frac{p(x)}{p'(x)} = \lim_{x \downarrow a} (x-a) \frac{\sum_{k=n_0}^{\infty} c_k(x-a)^{k-n_0}}{\sum_{k=n_0}^{\infty} k c_k(x-a)^{k-n_0}} \\ &= \frac{c_{n_0}}{n_0 c_{n_0}} \lim_{x \downarrow a} (x-a) = 0. \end{aligned}$$

In order to prove (f) we show that always

$$(50) \quad \liminf_{x \downarrow a} \frac{F(x)}{p(x)} = 0,$$

if  $p$  satisfies Condition 3.1. To show this, define the function  $G(x) := \log F(x)$  for  $x \in (a, b)$ . Then,  $G$  is increasing and continuously differentiable on  $(a, b)$  and satisfies  $G(a+) = -\infty$  and  $G(b-) = 0$ . If (50) did not hold, then

$$(51) \quad c := \limsup_{x \downarrow a} G'(x) = \limsup_{x \downarrow a} \frac{p'(x)}{p(x)} < +\infty.$$

Hence, choosing  $\delta > 0$  such that  $G'(x) \leq c + 1$  for all  $x \in (a, a + \delta]$  we would obtain

$$G(a+) - G(a + \delta) = -\lim_{x \downarrow a} \int_x^{a+\delta} G'(t)dt \geq -\lim_{x \downarrow a} (a + \delta - x)(c + 1) = -\delta(c + 1),$$

which would contradict  $G(a+) = -\infty$ . □

*Proof of Proposition 3.9.* With  $\tilde{h} = h - E[h(Z)]$ , since  $I = \eta \cdot p$ , we have for  $a < x < b$

$$|g_h(x)| = \frac{|\int_a^x \tilde{h}(t)p(t)dt|}{|p(x)\eta(x)|} = \frac{|\int_a^x \tilde{h}(t)p(t)dt|}{I(x)} \leq \|\tilde{h}\|_{\infty} \frac{F(x)}{I(x)}.$$

Let  $M : (a, b) \rightarrow \mathbb{R}$  be given by  $M(x) := \frac{F(x)}{I(x)}$ . By l'Hôpital's rule we have

$$\lim_{x \searrow a} M(x) = \lim_{x \searrow a} \frac{p(x)}{\gamma(x)p(x)} = \lim_{x \searrow a} \frac{1}{\gamma(x)} = \frac{1}{\lim_{x \searrow a} \gamma(x)}$$

which exists in  $[0, \infty)$  by Condition 3.2. Here, we used the convention  $\frac{1}{\infty} = 0$ . Moreover,

$$\lim_{x \nearrow b} M(x) = \frac{1}{\lim_{x \nearrow b} I(x)} = +\infty$$

again by Condition 3.2 and by Proposition 3.3. Furthermore, we have

$$(52) \quad M'(x) = \frac{p(x)I(x) - p(x)\gamma(x)F(x)}{I(x)^2} = \frac{p(x)}{I(x)^2} (I(x) - \gamma(x)F(x)) \geq 0$$

for each  $x \in (a, b)$  since by the positivity of  $p$  and because  $\gamma$  is decreasing

$$I(x) = \int_a^x \gamma(t)p(t)dt \geq \gamma(x) \int_a^x p(t)dt = \gamma(x)F(x).$$

Hence,  $M$  is increasing and, thus, for each  $x \in (a, m]$  we have

$$|g_h(x)| \leq \|\tilde{h}\|_\infty \frac{F(m)}{I(m)} = \frac{\|h - E[h(Z)]\|_\infty}{2I(m)}.$$

The same bound can be proved for  $x \in (m, b)$  by using the representation

$$g_h(x) = -\frac{1}{I(x)} \int_x^b (h(t) - E[h(Z)])p(t)dt$$

and the fact that also  $1 - F(m) = \frac{1}{2}$ . □

The following two lemmas, which are quite standard in Stein's method, will be needed for the proof of Proposition 3.13. For proofs we refer to [11], for instance.

**LEMMA 5.1.** *Suppose that  $p$  satisfies Condition 3.1 and that  $\int_a^b |x|p(x)dx < \infty$ . Then, for each  $x \in \overline{(a, b)}$  we have:*

- (a)  $\int_a^x F(t)dt = xF(x) - \int_a^x sp(s)ds$ ;
- (b)  $\int_x^b (1 - F(t))dt = \int_x^\infty sp(s)ds - x(1 - F(x))$ .

**LEMMA 5.2.** *Suppose that  $p$  satisfies Condition 3.1 and that  $E|Z| = \int_a^b |x|p(x)dx < \infty$ . Then, for each Lipschitz function  $h$ , the following assertions hold true:*

- (a) *For each  $y \in \mathbb{R}$  we have*  

$$h(y) - E[h(Z)] = \int_{-\infty}^y F(s)h'(s)ds - \int_y^\infty (1 - F(s))h'(s)ds.$$
- (b) *For each  $x \in \overline{(a, b)}$  we have*  

$$\int_a^x (h(y) - E[h(Z)])p(y)dy = -(1 - F(x)) \int_a^x F(s)h'(s)ds - F(x) \int_x^b (1 - F(s))h'(s)ds.$$

*Proof of Proposition 3.13.* First, we prove (a). Recall the representation

$$g_h(x) = \frac{1}{I(x)} \int_a^x (h(y) - E[h(Z)])p(y)dy.$$

By Lemmas 5.2 and 5.1 we thus obtain that

$$\begin{aligned}
& |I(x)g_h(x)| \\
&= \left| -(1-F(x)) \int_a^x F(s)h'(s)ds - F(x) \int_x^b (1-F(s))h'(s)ds \right| \\
&\leq \|h'\|_\infty \left( (1-F(x)) \int_a^x F(s)ds + F(x) \int_x^b (1-F(s))ds \right) \\
&= \|h'\|_\infty \left( (1-F(x)) \left( xF(x) - \int_a^x sp(s)ds \right) + F(x) \left( -x(1-F(x)) + \int_x^b sp(s)ds \right) \right) \\
&= \|h'\|_\infty \left( - \int_a^x sp(s)ds + F(x) \left( \int_a^x sp(s)ds + \int_x^b sp(s)ds \right) \right) \\
&= \|h'\|_\infty \left( F(x)E[Z] - \int_a^x yp(y)dy \right),
\end{aligned}$$

implying (a).

Now, we turn to the proof of (b). By Stein's equation (18) we obtain for  $x \in (a, b)$

$$(53) \quad g'_h(x) = \frac{1}{\eta(x)} \left( \tilde{h}(x) - \gamma(x)g_h(x) \right),$$

where we have again written  $\tilde{h} = h - E[h(Z)]$ . Using Lemma 5.2 again, we obtain

$$\begin{aligned}
(54) \quad g'_h(x) &= \frac{1}{\eta(x)} \left( \int_a^x F(s)h'(s)ds \left( 1 + \frac{\gamma(x)(1-F(x))}{\eta(x)p(x)} \right) \right. \\
&\quad \left. + \int_x^b (1-F(s))h'(s)ds \left( -1 + \frac{\gamma(x)F(x)}{\eta(x)p(x)} \right) \right) \\
&= \int_a^x F(s)h'(s)ds \left( \frac{\eta(x)p(x) + \gamma(x)(1-F(x))}{\eta(x)^2p(x)} \right) \\
&\quad + \int_x^b (1-F(s))h'(s)ds \left( \frac{-\eta(x)p(x) + \gamma(x)F(x)}{\eta(x)^2p(x)} \right).
\end{aligned}$$

Now, consider the functions  $H, G : \overline{(a, b)} \rightarrow \mathbb{R}$  with

$$\begin{aligned}
H(x) &= I(x) - \gamma(x)F(x) = \eta(x)p(x) - \gamma(x)F(x) \quad \text{and} \\
G(x) &= H(x) + \gamma(x) = \eta(x)p(x) + \gamma(x)(1-F(x)).
\end{aligned}$$

From (52) we already know that  $H$  is nonnegative on  $(a, b)$ . Similarly we prove the nonnegativity of  $G$  on  $(a, b)$ : Since  $p$  is positive and  $\gamma$  is decreasing, for  $x$  in  $(a, b)$  we have

$$\begin{aligned}
G(x) &= I(x) + \gamma(x)(1-F(x)) = - \int_x^b \gamma(t)p(t)dt + \gamma(x)(1-F(x)) \\
&\geq -\gamma(x)(1-F(x)) + \gamma(x)(1-F(x)) = 0.
\end{aligned}$$

By (54) we can thus bound

$$(55) \quad |g'_h(x)| \leq \|h'\|_\infty \left( \int_a^x F(s) ds \frac{G(x)}{\eta(x)^2 p(x)} + \int_x^b (1 - F(s)) ds \frac{H(x)}{\eta(x)^2 p(x)} \right),$$

which reduces to the bound asserted in (b). Optimality of the bound in (a) follows from choosing  $h(x) = x$  and observing that the above inequalities are in fact equalities, in this case. To see that also the bound in (b) is optimal, for given  $x \in (a, b)$  choose a 1-Lipschitz function  $h$  such that  $h'(s) = 1$  for all  $s \in (a, x)$  and  $h'(s) = -1$  for all  $s \in (x, b)$ . Then, from (54) and the nonnegativity of  $H$  and  $G$ , we see that equality holds in (55). □

*Proof of Corollary 3.15.* Claim (a) follows from Proposition 3.13 (a) and the observation that in this case we have

$$I(x) = \int_a^x \gamma(y) p(y) dy = c \int_a^x (E[Z] - y) p(y) dy.$$

Part (b) follows from Proposition 3.13 (b) and Lemma 5.1 by observing that in this case

$$\begin{aligned} H(x) &= I(x) - \gamma(x)F(x) = c \left( \int_a^x (E[Z] - t) p(t) dt - (E[Z] - x) F(x) \right) \\ &= c \left( E[Z] F(x) - \int_a^x t p(t) dt - E[Z] F(x) + x F(x) \right) \\ &= c \int_a^x F(s) ds \end{aligned}$$

and, similarly,  $G(x) = c \int_x^b (1 - F(s)) ds$ . □

*Proof of Proposition 3.12.* We first prove necessity. Let  $f$  be given as in the proposition. First we show that  $E|\gamma(Z)f(Z)| < \infty$ . We have

$$\begin{aligned}
E|\gamma(Z)f(Z)| &= \int_a^{x_0} \gamma(x)|f(x)|p(x)dx - \int_{x_0}^b \gamma(x)|f(x)|p(x)dx \\
&= \int_a^{x_0} \gamma(x)p(x) \left| \int_x^{x_0} f'(t)dt - f(x_0) \right| dx - \int_{x_0}^b \gamma(x)p(x) \left| \int_{x_0}^x f'(t)dt + f(x_0) \right| dx \\
&\leq |f(x_0)| \left( \int_a^{x_0} \gamma(x)p(x)dx - \int_{x_0}^b \gamma(x)p(x)dx \right) + \int_a^{x_0} \gamma(x)p(x) \int_x^{x_0} |f'(t)|dt \\
&\quad - \int_{x_0}^b \gamma(x)p(x) \int_{x_0}^x |f'(t)|dt \\
&= |f(x_0)|E|\gamma(Z)| + \int_a^{x_0} |f'(t)| \int_a^t \gamma(x)p(x)dxdt - \int_{x_0}^b |f'(t)| \int_t^b \gamma(x)p(x)dxdt \\
&= |f(x_0)|E|\gamma(Z)| + \int_a^b |f'(t)|I(t)dt < \infty.
\end{aligned}$$

Repeating essentially the same calculation without absolute value signs and using  $E[\gamma(Z)] = 0$  yields

$$E[\eta(Z)f'(Z)] = -E[\gamma(Z)f(Z)].$$

To prove sufficiency it is clearly enough to show that

$$E[h(X)] = E[h(Z)]$$

holds for each bounded and continuous function  $h$ . Let  $g_h$  be the standard solution of the Stein equation (18) corresponding to  $h$ . Then, from Proposition 3.9 we know that  $\|g_h\|_\infty < \infty$ . Also,  $g_h$  is continuous on  $(\overline{a}, \overline{b})$  and continuously differentiable on each compact subinterval of  $(\overline{a}, \overline{b})$ . Furthermore, since  $I(x) = \eta(x)p(x)$  and  $g_h$  solves (18) we have

$$|g'_h(x)I(x) = p(x)|\tilde{h}(x) - \gamma(x)g_h(x)| \leq p(x)|\tilde{h}(x)| + |\gamma(x)g_h(x)|$$

and, hence,

$$\int_a^b |g'_h(x)I(x)dx| \leq 2\|h\|_\infty + \|g_h\|_\infty E|\gamma(Z)| < \infty.$$

By the hypothesis of Proposition 3.12 we can thus conclude that

$$0 = E[\eta(X)g'_h(X) + \gamma(X)g_h(X)] = E[h(X)] - E[h(Z)],$$

as desired. □

*Proof of Proposition 3.17.* We first show that  $E[h_2(\tilde{Z})]$  exists. Since  $\tilde{Z}$  has density proportional to  $\eta p$  by (37) and by (36), existence follows, if we can show that

$$(56) \quad \int_a^b |\gamma'(x)| |g_h(x)\eta(x)p(x)|dx < \infty \quad \text{and} \quad \int_a^b |h'(x)| \eta(x)p(x)dx < \infty.$$

To show finiteness of the first integral in (56) note that since  $\gamma$  is decreasing, by Fubini's theorem

$$\begin{aligned}
& \int_a^{x_0} |\gamma'(x)| |g_h(x) \eta(x) p(x)| dx = - \int_a^{x_0} \gamma'(x) \left| \int_a^x \tilde{h}(t) p(t) dt \right| dx \\
& \leq - \int_a^{x_0} \gamma'(x) \int_a^x |\tilde{h}(t)| p(t) dt dx = \int_a^{x_0} |\tilde{h}(t)| p(t) \int_t^{x_0} (-\gamma'(x)) dx dt \\
& = \int_a^{x_0} |\tilde{h}(t)| p(t) (\gamma(t) - \gamma(x_0)) dt = \int_a^{x_0} |\tilde{h}(t)| p(t) \gamma(t) dt \\
(57) \quad & \leq \|h'\|_\infty \int_a^{x_0} |t - Z| p(t) \gamma(t) dt \leq \|h'\|_\infty \left( E|Z| E|\gamma(Z)| + E|Z\gamma(Z)| \right) < \infty,
\end{aligned}$$

since  $h$  is Lipschitz. Similarly, one shows that

$$\int_{x_0}^b |\gamma'(x)| |g_h(x) \eta(x) p(x)| dx < \infty.$$

Since  $h'$  is bounded, to show that the second integral in (56) is finite, it suffices to prove that

$$(58) \quad \int_a^b \eta(x) p(x) dx < \infty.$$

We have

$$\begin{aligned}
\int_a^{x_0} \eta(x) p(x) dx &= \int_a^{x_0} \int_a^x \gamma(t) p(t) dt dx \\
&= \int_a^{x_0} \gamma(t) p(t) \int_t^{x_0} dx dt = \int_a^{x_0} (x_0 - t) \gamma(t) p(t) dt < \infty,
\end{aligned}$$

since  $E|\gamma(Z)| < \infty$  and  $E|Z\gamma(Z)| < \infty$  and similarly one shows that

$$\int_{x_0}^b \eta(x) p(x) dx < \infty.$$

Hence, (58) holds and  $E[h_2(\tilde{Z})]$  exists. Now, we prove that  $E[h_2(\tilde{Z})] = 0$ . From (36) and (37) we see that this amounts to proving

$$(59) \quad \int_a^b h'(x) \eta(x) p(x) dx = \int_a^b \gamma'(x) g_h(x) \eta(x) p(x) dx.$$

Using  $\eta p = I$ ,  $I' = \gamma p$  and  $I(a+) = I(b-) = 0$ , from Fubini's theorem we obtain that the left hand side of (59) equals

$$\begin{aligned}
 & \int_a^{x_0} h'(x)I(x)dx + \int_{x_0}^b h'(x)I(x)dx \\
 &= \int_a^{x_0} h'(x) \int_a^x I'(t)dt dx - \int_{x_0}^b h'(x) \int_x^b I'(t)dt dx \\
 &= \int_a^{x_0} I'(t) \int_t^{x_0} h'(x)dx dt - \int_{x_0}^b I'(t) \int_{x_0}^t h'(x)dx dt \\
 &= \int_a^{x_0} I'(t)(h(x_0) - h(t))dt - \int_{x_0}^b I'(t)(h(t) - h(x_0))dt \\
 (60) \quad &= \int_a^b \gamma(t)p(t)(h(x_0) - h(t))dt = - \int_a^b h(t)\gamma(t)p(t)dt.
 \end{aligned}$$

Similarly, using  $\gamma(x_0) = 0$ , the definition of  $g_h$  in (29) and Fubini's theorem again, we have that the right hand side of (59) equals

$$\begin{aligned}
 & \int_a^{x_0} \gamma'(x)g_h(x)\eta(x)p(x)dx + \int_{x_0}^b \gamma'(x)g_h(x)\eta(x)p(x)dx \\
 &= \int_a^{x_0} \gamma'(x) \int_a^x \tilde{h}(t)p(t)dt dx - \int_{x_0}^b \gamma'(x) \int_t^b \tilde{h}(t)p(t)dt dx \\
 &= \int_a^{x_0} \tilde{h}(t)p(t) \int_t^{x_0} \gamma'(x)dx dt - \int_{x_0}^b \tilde{h}(t)p(t) \int_{x_0}^t \gamma'(x)dx dt \\
 (61) \quad &= - \int_a^b \tilde{h}(t)p(t)\gamma(t)dt = - \int_a^b h(t)\gamma(t)p(t)dt,
 \end{aligned}$$

where we have used  $E[\gamma(Z)] = 0$  for the last equality. Thus, from (60) and (61) we conclude that (59) holds.

Thus, the standard solution  $f = f_{h_2}$  to (38) is well-defined and given by

$$f(x) = \frac{1}{\eta(x)^2 p(x)} \int_a^x h_2(t)\eta(t)p(t)dt = \frac{-1}{\eta(x)^2 p(x)} \int_x^b h_2(t)\eta(t)p(t)dt, \quad a < x < b.$$

Hence,

$$(63) \quad \lim_{x \downarrow a} \eta(x)^2 p(x) f(x) = 0 = \lim_{x \uparrow b} \eta(x)^2 p(x) f(x)$$

by dominated convergence. Furthermore, since  $g'_h$  is also a solution to (38) and the solutions to the corresponding homogeneous equation are exactly the constant multiples of  $\eta^{-2}p^{-1}$ , there is a constant  $c \in \mathbb{R}$  such that

$$(64) \quad g'_h(x) = f(x) + \frac{c}{\eta(x)^2 p(x)}, \quad x \in (a, b).$$

Now, first suppose that  $a > -\infty$ . Since  $g_h$  solves the Stein equation (18) we know that

$$(65) \quad \eta(x)^2 p(x) g'_h(x) = I(x)(\tilde{h}(x) - \gamma(x)g_h(x)).$$

As  $x \downarrow a > -\infty$ , by (30), the term in brackets converges to  $\tilde{h}(a) - \gamma(a+) \frac{\tilde{h}(a)}{\gamma(a+)} = 0$  and since  $\lim_{x \rightarrow a} I(x) = 0$  we conclude from (65) that also

$$(66) \quad \lim_{x \downarrow a} \eta(x)^2 p(x) g'_h(x) = 0.$$

Hence, from (63), (66) and (64) we conclude that  $g'_h = f$  is the standard solution to (38). Similarly, one obtains this result if  $b < \infty$ . Finally assume that  $g'_h$  is bounded. Since  $\lim_{x \downarrow a} \eta(x)^2 p(x) = 0$  we conclude from (64) and (63) that

$$0 = \lim_{x \downarrow a} \eta(x)^2 p(x) g'_h(x) = \lim_{x \downarrow a} \eta(x)^2 p(x) f(x) + c = c.$$

□

*Proof of Proposition 3.19.* Let  $x_0$  be defined as above and define the function  $G : J \rightarrow \mathbb{R}$  by  $G(x) := \int_{x_0}^x f(y) dy$ . Then, by Taylor's formula, for each  $x, x' \in I$  we have

$$\begin{aligned} G(x') - G(x) &= G'(x)(x' - x) + \int_x^{x'} (x' - t) G''(t) dt \\ &= f(x)(x' - x) + \int_x^{x'} (x' - t) f'(t) dt \\ &= f(x)(x' - x) + (x' - x)^2 \int_0^1 (1 - s) f'(x + s(x' - x)) ds. \end{aligned}$$

Hence, by distributional equality, we obtain

$$\begin{aligned} 0 &= E[G(W')] - E[G(W)] \\ &= E[f(W)(W' - W)] + E\left[(W' - W)^2 \int_0^1 (1 - s) f'((W + s(W' - W))) ds\right] \\ &= E[f(W)E[W' - W|W]] + E\left[(W' - W)^2 \int_0^1 (1 - s) f'((W + s(W' - W))) ds\right] \\ &= \lambda E[f(W)\gamma(W)] + \lambda E[f(W)R] \\ &\quad + E\left[(W' - W)^2 \int_0^1 (1 - s) f'((W + s(W' - W))) ds\right], \end{aligned}$$

yielding

$$(67) \quad E[f(W)\gamma(W)] = -\frac{1}{\lambda} E\left[(W' - W)^2 \int_0^1 (1 - s) f'((W + s(W' - W))) ds\right] - E[f(W)R].$$

This immediately implies the identity

$$\begin{aligned} &E[\eta(W)f'(W) + \gamma(W)f(W)] \\ &= E\left[f'(W)(\eta(W) - \frac{1}{2\lambda}(W' - W)^2)\right] \\ &\quad + \frac{1}{\lambda} E\left[(W' - W)^2 \int_0^1 (1 - s) \left(f'(W) - f'((W + s(W' - W)))\right) ds\right] \\ (68) \quad &- E[f(W)R]. \end{aligned}$$



Observing that

$$\left| f'(W) - f'((W + s(W' - W))) \right| \leq \|f''\|_\infty s|W' - W|$$

and  $\int_0^1 s(1-s)ds = \frac{1}{6}$  the bound (39) now easily follows from (68) and the properties of  $f$ . □

*Proof of Proposition 4.2.* Claim (a) immediately follows from Proposition 3.9. Similarly, the first part of claim (b) immediately follows from Corollary 3.15 (a). For the second part of (b) we note that by Corollary 3.15 (b) we have for  $x \in (0, 1)$ :

$$(69) \quad |g'_h(x)| \leq 2(a+b)\|h'\|_\infty \frac{\int_0^x F(t)dt \int_x^1 (1-F(s))ds}{x^2(1-x)^2p(x)}, \quad x \in (0, 1).$$

Since  $F$  is increasing and  $1-F$  is decreasing, we have

$$\int_0^x F(t)dt \leq xF(x) \quad \text{and} \quad \int_x^1 (1-F(s))ds \leq (1-F(x))(1-x)$$

for each  $x \in [0, 1]$ . Plugging this into (69) yields

$$(70) \quad |g'_h(x)| \leq 2(a+b)S(x)\|h'\|_\infty,$$

where

$$S(x) = \frac{F(x)(1-F(x))}{\eta(x)p(x)} = \frac{1}{B(a,b)} \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt \int_x^1 s^{a-1}(1-s)^{b-1}ds}{x^a(1-x)^b}.$$

By de l'Hôpital's rule, one can easily show that  $S(0+) = a^{-1}$  and  $S(1-) = b^{-1}$ . Thus, it suffices to bound  $\|S\|_\infty$ . For general  $a, b > 0$  we write

$$(71) \quad S(x) = \frac{1}{B(a,b)} f_1(x) f_2(x),$$

where

$$f_1(x) := f_1(x; a, b) := \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{x^a} \quad \text{and} \\ f_2(x) := f_2(x; a, b) := \frac{\int_x^1 t^{a-1}(1-t)^{b-1}dt}{(1-x)^b}.$$

For  $a \neq b$  we bound the functions  $f_1$  and  $f_2$  separately. By de l'Hôpital's rule we have

$$\lim_{x \downarrow 0} f_1(x) = \lim_{x \downarrow 0} \frac{x^{a-1}(1-x)^{b-1}}{ax^{a-1}} = \frac{1}{a}.$$

Also, note that

$$f'_1(x) = \frac{x^a(1-x)^{b-1} - a \int_0^x t^{a-1}(1-t)^{b-1}dt}{x^{a+1}} =: \frac{N_1(x)}{x^{a+1}}.$$

We have  $N_1(0+) = 0$  and

$$N'_1(x) = (1-b)x^a(1-x)^{b-2} \begin{cases} \geq 0 \quad \forall x \in (0, 1), & b \leq 1 \\ < 0 \quad \forall x \in (0, 1), & b > 1. \end{cases}$$

This implies that  $N_1$  is nonnegative and, hence,  $f_1$  is increasing for  $b \leq 1$  and that  $N_1$  is nonpositive and, hence,  $f_1$  is decreasing for  $b > 1$ . Thus,

$$(72) \quad \|f_1\|_\infty = \begin{cases} f_1(1-) = B(a, b), & b \leq 1 \\ f_1(0+) = a^{-1}, & b > 1. \end{cases}$$

Since  $f_2(x; a, b) = f_1(1 - x; b, a)$  we have

$$(73) \quad \|f_2\|_\infty = \begin{cases} B(b, a) = B(a, b), & a \leq 1 \\ b^{-1}, & a > 1. \end{cases}$$

Thus, from (70), (71), (72) and (73) we have

$$\|g'_h\|_\infty \leq C(a, b)\|h'\|_\infty,$$

where  $C(a, b)$  is given by (47) and (48). In the case  $a = b$  we can provide better bounds. First note that in this case the Beta distribution  $Beta(a, a)$  is symmetric with respect to  $1/2$ . This easily implies that

$$S(1/2 - x) = S(1/2 + x)$$

holds for each  $0 \leq x \leq 1/2$ . Thus it suffices to bound  $S$  on  $[1/2, 1)$ . Note that

$$S(1/2) = \frac{F(1/2)(1 - F(1/2))}{1/2(1 - 1/2)p(1/2)} = \frac{1}{p(1/2)} = 2^{2a-2}B(a, a)$$

and

$$S(1-) = F(1) \lim_{x \uparrow 1} \frac{1 - F(x)}{\eta(x)p(x)} = \lim_{x \uparrow 1} \frac{-p(x)}{\gamma(x)p(x)} = \frac{-1}{\gamma(1)} = \frac{1}{b} = \frac{1}{a}.$$

For  $x \in (0, 1)$  we have

$$(74) \quad S'(x) = \frac{T(x)}{\eta(x)^2 p(x)},$$

where

$$(75) \quad T(x) = \eta(x)p(x)(1 - 2F(x)) - F(x)(1 - F(x))\gamma(x).$$

Thus,  $S$  is increasing (decreasing) on  $[1/2, 1)$ , if and only if  $T$  is nonnegative (non-positive) there. In the case  $a = b$  we have  $\gamma(x) = a(1 - 2x)$  and, hence,  $\gamma(1/2) = F(1/2) = 0$ . Thus, recalling that  $I(1) = (\eta p)(1-) = 0$  we have

$$(76) \quad T(1/2) = T(1-) = 0.$$

By (76) the nonnegativity (nonpositivity) of  $T$  on  $[1/2, 1)$  follows, if  $T(y) \geq 0$  ( $\leq 0$ ) for every locally extremal point  $y \in (1/2, 1)$ . We have

$$(77) \quad T'(x) = -2\eta(x)p(x)^2 - \gamma'(x)F(x)(1 - F(x))$$

and, hence, if  $y \in (1/2, 1)$  is a locally extremal point of  $T$ , we have  $T'(y) = 0$  and

$$(78) \quad \begin{aligned} T(y) &= \eta(y)p(y) \left( 1 - 2F(y) + 2 \frac{\gamma(y)}{\gamma'(y)} p(y) \right) \\ &= \eta(y)p(y) \left( 1 - 2F(y) + (2y - 1)p(y) \right). \end{aligned}$$

Now, for  $x \in [1/2, 1)$ , consider the function

$$U(x) = 1 - 2F(x) + (2x - 1)p(x)$$

and note that  $U(1/2) = 0$ . For  $1/2 \leq x < 1$  we have

$$U'(x) = (2x - 1)p'(x) = p(x)(2x - 1)\psi(x) = \frac{p(x)(1 - 2x)^2}{x(1 - x)}(1 - a)$$

and, hence,  $U$  is increasing for  $a \leq 1$  and is decreasing for  $a \geq 1$ . Since  $U(1/2) = 0$  it thus follows from (78) that if  $y \in (1/2, 1)$  is a locally extremal point of  $T$ , then  $T(y)$  is nonnegative for  $a < 1$  and nonpositive for  $a \geq 1$ . From (74) and (76) it thus follows that  $S$  is decreasing on  $[1/2, 1)$  if  $a \geq 1$  and increasing if  $a < 1$ . Hence, we can conclude that

$$(79) \quad \|S\|_\infty = \sup_{x \in [1/2, 1)} S(x) = \begin{cases} S(1-) = a^{-1}, & 0 < a < 1 \\ S(1/2) = p(1/2)^{-1}, & a \geq 1. \end{cases}$$

Note that by the duplication formula for the Gamma function we have

$$\begin{aligned} p(1/2) &= B(a, a)^{-1} \left(\frac{1}{2}\right)^{2a-2} = \left(\frac{1}{2}\right)^{2a-2} \frac{\Gamma(2a)}{\Gamma(a)^2} \\ &= \frac{2^{2a-1} \Gamma(a + 1/2) \Gamma(a)}{\sqrt{\pi} \Gamma(a)^2 2^{2a-2}} = \frac{2 \Gamma(a + 1/2)}{\sqrt{\pi} \Gamma(a)}. \end{aligned}$$

Hence, by (70) and (79) this implies

$$\|g'_h\|_\infty \leq 4a \|S\|_\infty \|h'\|_\infty = \|h'\|_\infty \begin{cases} 4, & 0 < a < 1 \\ \frac{2a\sqrt{\pi}\Gamma(a)}{\Gamma(a+1/2)}, & a \geq 1. \end{cases}$$

Now, we turn to the proof of (c). From Proposition 3.17 we know that  $g'_h$  is the standard solution to the Stein equation

$$x(1-x)f'(x) + (a+1-(a+b+2)x)f(x) = h_2(x) = h'(x) + (a+b)g_h(x)$$

corresponding to the distribution  $Beta(a+1, b+1)$ . Thus, since  $h_2$  is Lipschitz by part (b), applying (b) for  $Beta(a+1, b+1)$  and for  $Beta(a, b)$  yields

$$\begin{aligned} \|g''_h\|_\infty &\leq C(a+1, b+1) \|h'_2\|_\infty \leq C(a+1, b+1) (\|h''\|_\infty + (a+b) \|g'_h\|_\infty) \\ (80) \quad &\leq C(a+1, b+1) \|h''\|_\infty + C(a+1, b+1) C(a, b) \|h'\|_\infty, \end{aligned}$$

as claimed. The proof of (d) is very similar to the proof of (c) which is why we only give a sketch. Defining  $h_1 := \tilde{h}$  and for  $k = 2, \dots, m$

$$h_k(x) = x(1-x)g_h^{(k)}(x) + (a+k-1-(a+b+2k-2)x)g_h^{(k-1)}(x),$$

one can see by induction that for all  $k = 2, \dots, m$

$$h_k = h^{(k-1)} + (k-1)(a+b+k-2)g_h^{(k-2)}.$$

Hence, by (b) and from Proposition 3.17 similarly to (80) we can prove that

$$\begin{aligned} \|g_h^{(m)}\|_\infty &\leq C(a+m-1, b+m-1) \|h'_m\|_\infty \\ &= C(a+m-1, b+m-1) \|h^{(m)} + (m-1)(a+b+m-2)g_h^{(m-1)}\|_\infty \\ &\leq C(a+m-1, b+m-1) \left( \|h^{(m)}\|_\infty + (m-1)(a+b+m-2) \|g_h^{(m-1)}\|_\infty \right). \end{aligned}$$

The bound now follows from an easy induction on  $m$ .

□

## REFERENCES

- [1] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.
- [2] A. D. Barbour. Stein’s method and Poisson process convergence. *J. Appl. Probab.*, (Special Vol. 25A):175–184, 1988. A celebration of applied probability.
- [3] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publications.
- [4] S. Chatterjee, J. Fulman, and A. Röllin. Exponential approximation by Stein’s method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat.*, 8:197–223, 2011.
- [5] S. Chatterjee and Q.-M. Shao. Nonnormal approximation by Stein’s method of exchangeable pairs with application to the Curie-Weiss model. *Ann. Appl. Probab.*, 21(2):464–483, 2011.
- [6] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probab.*, 3(3):534–545, 1975.
- [7] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein’s method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.
- [8] L. H. Y. Chen and Q.-M. Shao. Stein’s method for normal approximation. In *An introduction to Stein’s method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–59. Singapore Univ. Press, Singapore, 2005.
- [9] F. Daly. Upper bounds for Stein-type operators. *Electron. J. Probab.*, 13:no. 20, 566–587, 2008.
- [10] C. Döbler. New developments in Stein’s method with applications. 2012. (Ph.D.)-Thesis Ruhr-Universität Bochum.
- [11] C. Döbler. Stein’s method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the Polya urn model. [arXiv:1207.0533](#), 2012.
- [12] R. Eden and J. Viquez. Nourdin-Peccati analysis on Wiener and Wiener-Poisson space for general distributions. *Stochastic Process. Appl.*, 125(1):182–216, 2015.
- [13] P. Eichelsbacher and M. Löwe. Stein’s method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.*, 15:no. 30, 962–988, 2010.
- [14] J. Fulman and N. Ross. Exponential approximation and Stein’s method of exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(1):1–13, 2013.
- [15] R. E. Gaunt. Rates of convergence in normal approximation under moment conditions via new bounds on solutions of the stein equation. *to appear in J. Theoret. Probab.*, 2014.
- [16] R. E. Gaunt. Variance-gamma approximation via stein’s method. *Electron. J. Probab.*, 19, 2014.
- [17] R. E. Gaunt, A. Pickett, and G. Reinert. Chi-square approximation by Stein’s method with application to Pearson’s statistic. [arXiv:1507.01707](#), 2015.
- [18] L. Goldstein and G. Reinert. Stein’s method for the beta distribution and the Pólya-Eggenberger urn. *J. Appl. Probab.*, 50(4):1187–1205, 2013.
- [19] L. Goldstein and Y. Rinott. Multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.
- [20] S. Kusuoka and C. A. Tudor. Stein’s method for invariant measures of diffusions via Malliavin calculus. *Stochastic Process. Appl.*, 122(4):1627–1651, 2012.
- [21] C. Ley, G. Reinert, and Y. Swan. Approximate computation of expectations: a canonical Stein operator. [arXiv:1408.2998](#), 2014.
- [22] H. M. Luk. *Stein’s method for the Gamma distribution and related statistical applications*. ProQuest LLC, Ann Arbor, MI, 1994. Thesis (Ph.D.)–University of Southern California.
- [23] I. Nourdin and G. Peccati. Stein’s method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009.
- [24] I. Nourdin and F. G. Viens. Density formula and concentration inequalities with Malliavin calculus. *Electron. J. Probab.*, 14:no. 78, 2287–2309, 2009.
- [25] E. A. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011.
- [26] E. A. Peköz, A. Röllin, and N. Ross. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.*, 23(3):1188–1218, 2013.

- [27] J. Pike and H. Ren. Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 11(1):571–587, 2014.
- [28] G. Reinert. Three general approaches to Stein's method. In *An introduction to Stein's method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 183–221. Singapore Univ. Press, Singapore, 2005.
- [29] A. Röllin. A note on the exchangeability condition in Stein's method. *Statist. Probab. Lett.*, 78(13):1800–1806, 2008.
- [30] W. Schoutens. Orthogonal polynomials in Stein's method. *J. Math. Anal. Appl.*, 253(2):515–531, 2001.
- [31] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.
- [32] C. Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.